



**HAL**  
open science

## Local social cohesion and economic outcomes in the city

Camille Hémet

► **To cite this version:**

Camille Hémet. Local social cohesion and economic outcomes in the city. Economics and Finance. Université Paris Sud, 2019. <tel-05010723>

**HAL Id: tel-05010723**

**<https://hal.science/tel-05010723v1>**

Submitted on 28 Mar 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

UNIVERSITÉ PARIS-SUD  
Ecole Doctorale Sciences de l'Homme et de la Société n°578  
FACULTÉ JEAN MONNET

**Thèse d'Habilitation à Diriger des Recherches**

*Présentée et soutenue publiquement par*

**Camille HÉMET**

*le 27 novembre 2019*

---

**LOCAL SOCIAL COHESION AND ECONOMIC  
OUTCOMES IN THE CITY**

---

JURY

<b>M. Pierre-Philippe COMBES</b> <i>Directeur de recherche au CNRS, GATE-LSE</i>	Rapporteur
<b>Mme. Gabrielle FACK</b> <i>Professeure, Université Paris Dauphine</i>	Rapporteur
<b>M. Laurent GOBILLON</b> <i>Directeur de recherche au CNRS, PSE</i>	Examineur
<b>Mme. Miren LAFOURCADE</b> <i>Professeure, Université Paris-Saclay</i>	Tutrice
<b>M. Florian MAYNERIS</b> <i>Professeur, Université du Québec à Montréal</i>	Rapporteur





*A Maxime, Lili et Mathis*



## Remerciements

29/10/2019, 15h15. Il est grand temps. Grand temps de penser à conclure ce mémoire par ces quelques mots de remerciements.

Je tiens tout d'abord à exprimer ma gratitude à ma tutrice Miren Lafourcade pour sa patience, son temps passé à relire ce mémoire et ses conseils avisés. Merci de m'avoir embarquée dans cette aventure, et d'y avoir cru plus que moi !

J'exprime ensuite ma sincère reconnaissance à Gabrielle Fack, Pierre-Philippe Combes et Florian Mayneris qui ont accepté d'être rapporteurs de ce mémoire, ainsi qu'à Laurent Gobillon qui fait partie de ce jury, dont les commentaires et conseils en cette occasion me seront précieux pour la suite.

Mes travaux de recherche passés n'auraient pas vu le jour sans la collaboration de co-auteurs hors-pair. Je remercie donc, par ordre d'apparition, Yann Algan, David Laitin, Elisabet Viladecans-Marsal, Miquel-Àngel Garcia-López et Clément Malgouyres. Je remercie également mes co-auteurs plus récents, sans qui mes travaux en cours ne seraient pas en cours : Nina Guyon, Arnaud Philippe, Manon Garrouste, Matthijs Oosterveen, Arnaud Chevalier, Bastien Michel, Paul Brandily et Clément Malgouyres. Que nos travaux puissent nous permettre de mieux comprendre les liens entre cohésion social et économie au niveau local.

Mes pensées se tournent à présent vers l'ensemble des collègues et des personnels administratifs de l'Institut d'Economie de Barcelone où j'ai effectué mon post-doctorat, de l'Ecole d'Economie de Paris et de l'ENS où je fais mes armes depuis maintenant quatre ans. Mes débuts de chercheuse n'auront été que facilités par la richesse de ces institutions, la qualité des interactions entre chercheurs, et l'aide précieuse des équipes administratives. Pour son assistance attentive et efficace au cours de ces derniers mois, je remercie particulièrement Laurence Vincent, qui m'a permis de libérer un temps précieux pour la réalisation de ce mémoire.

Sur une note plus personnelle, il va sans dire que je n'aurais pas su trouver les ressources nécessaires pour mener ce projet à bien sans le soutien infaillible de mon mari.

# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	What is (or not) social cohesion? How can we measure it? . . . . .	4
1.2	Research overview . . . . .	6
1.2.1	Local social cohesion and economic conditions . . . . .	6
1.2.2	Urban environment and socio-economic outcomes . . . . .	11
<b>2</b>	<b>Local social cohesion and economic conditions</b>	<b>13</b>
2.1	On the influence of social interactions on social and economic outcomes . . . . .	13
2.1.1	Interactions and housing conditions in a context of diversity of origins . . . . .	14
2.1.2	Social interactions and employment - <i>Working and published papers</i> . . . . .	15
2.1.3	Social interactions and education - <i>Recent ongoing projects</i> . . . . .	22
2.2	On the influence of economic conditions on social cohesion . . . . .	24
2.2.1	Spatial economic disparities and crime victimization . . . . .	24
2.2.2	Economic conditions, local public finance and electoral outcomes - <i>Future project</i> . . . . .	27
2.3	Institutional environment, economic conditions and social cohesion . . . . .	27
<b>3</b>	<b>Urban environment and socio-economic outcomes</b>	<b>30</b>
3.1	From transportation infrastructure to metropolitan employment . . . . .	30
3.2	From urban amenities and infrastructure to social cohesion . . . . .	35
<b>4</b>	<b>Research-related activities</b>	<b>37</b>
4.1	Masters' students supervision and related activities . . . . .	37
4.2	Experience related to PhD students supervision . . . . .	38
4.3	Contribution to scientific projects and events . . . . .	39
4.4	Other administrative and collective responsibilities . . . . .	39
<b>5</b>	<b>Bibliography</b>	<b>40</b>

<b>A Curriculum Vitae</b>	<b>48</b>
<b>B List of Published and Unpublished Works</b>	<b>54</b>
<b>C Presentation of research</b>	<b>55</b>
<b>D Research Articles</b>	<b>56</b>
D.1 The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation . . . . .	57
D.2 Diversity and Employment Prospects: Neighbors Matter! . . . . .	113
D.3 Ethnic Networks and the Informal Labor Market . . . . .	148
D.4 The Strength and Use of Local Referral Networks: Evidence from France . . . . .	183
D.5 Unemployment and Crime Victimization: A Local Approach . . . . .	201
D.6 Custodial vs. Non-Custodial Punishments: Measuring the Impact of an Anticipated Reform .	236
D.7 How does transportation shape intrametropolitan growth? An answer from the Regional Express Rail . . . . .	287
D.8 Next train to the polycentric city: The effect of railroads on subcenter formation . . . . .	311

# 1 Introduction

In 2015, as much as 80.4 percent of high income countries' population lived in an urban area. Still, the United Nations expect this share to increase up to 86.7 percent by 2050. Ultimately, this trend implies that individuals will live in denser and denser cities, increasing the potential for social and economic interactions. Such interactions can be positive and lead to productive outcomes (e.g. exchange of information reducing labor market frictions), but they can also be detrimental to the community (e.g. criminal interactions). It is thus crucial to understand what are the conditions for urban settlements to evolve toward socially and economically sustainable societies. In particular, the organization of the urban space (location and quality of buildings, transportation infrastructures, public parks...) may shape the nature of social encounters and hence influence the social and economic sphere. The purpose of my research is precisely to investigate whether and how urban infrastructure and amenities can foster social cohesion and lead to economically and socially sustainable cities, in the French context.

Understanding how cities form, how they are organized and how their structure can enhance local growth and economic development are key issues that have been at the center of research in urban economics for several decades. Following [Henderson \(1974\)](#), urban economics has sought to explain the existence and size of cities as resulting from two opposite forces: agglomeration effects pertaining to the productive sphere, such as labor market pooling, input sharing, and knowledge spillovers (see [Rosenthal and Strange \(2004\)](#) for a review), and negative congestion effects reflecting on consumers, such as longer commuting, increased pollution or higher rents ([Combes et al., 2018](#)). However, as pointed out by [Edward L. Glaeser and Saiz \(2001\)](#), little attention is paid to the consumption-related agglomeration forces, i.e. the extent to which urban density provides attractive features to consumers.

Yet, understanding what determines individuals' willingness and ability to live close to each other is crucial to figure out the conditions for sustainable human settlements. Indeed, unappealing locations are likely

to be characterized by high residential turnover and strong negative social sorting, impeding the creation and enforcement of social norms and leading to adverse economic conditions and socially fragmented communities. On the other hand, some consumption amenities reflecting high quality of life may overcome the dispersion forces related to congestion and make cities more attractive, as suggested by [Rappaport \(2009\)](#). Such amenities include for instance the access to a rich variety of local consumer goods and services, the possibility of a large and diverse set of social interactions, the access to public services, and the (time) proximity to (spatially) distant locations made possible by efficient transportation infrastructure. Cities offering such amenities could instead be characterized by high residential stability, providing ground for more integrated human settlement and more cohesive societies through the creation and enforcement of common social norms.

This is the type of mechanisms and relationships that my research aims at uncovering by asking the following question: how can urban amenities – ranging from local public goods and services to transportation infrastructure – promote social cohesion and eventually lead to economically and socially sustainable communities? Given that we are living in an increasingly urbanized world, it is critical to figure out the conditions making cities socially cohesive places where people share trust relationships, common norms and a sense of belonging to a community, not only within but also across social or ethnic groups. In turn, socially cohesive communities characterized by higher levels of generalized trust promote cooperation and open the way towards economic security and stability (see [Zak and Knack, 2001](#); [Francois and Zbojnik, 2005](#); [East-erly et al., 2006](#); [Algan and Cahuc, 2010](#), among others).

Overall, my research lies at the crossroad of two economic literatures: it first and foremost relates to the branch of urban economics focusing on city infrastructure and amenities but it is also connected to recent research on the growth effect of social cohesion and trust. Yet, the role of a city's "physical" environment (consumption amenities, equipment, access to public services, transportation infrastructure) on social cohesion remains essentially absent from both literatures. My research contributes to fill this gap by explicitly relating the urban environment to social cohesion.

As mentioned above, the role of consumption amenities on cities' attractiveness was initially left aside by the urban economics literature. Following [Edward L. Glaeser and Saiz \(2001\)](#), several recent papers have shown that natural, aesthetic or consumption amenities are important determinants of residential choices

(Rosenthal and Ross, 2015), thus affecting housing prices, land rents and the spatial income distribution of cities (Rappaport, 2008; Albouy, 2016; Koster et al., 2016; Gaigné et al., 2017). Compared to this literature, my current research proposes to zoom further into the picture and understand how amenities shape cities' social fabric. To put it differently, the purpose here is not only to understand how amenities and infrastructure attract residents (and how it sorts them across space), but mostly to investigate how they facilitate trustful social encounters (within or across social or ethnic groups). Moreover, although the effect of amenities and infrastructure on employment and growth has been revealed in previous studies (see for instance Duranton and Turner, 2012), the underlying mechanisms remain largely unexplored. My research sheds some light to this respect as well, by exploring the social cohesion channel thanks to its very local approach.

On the other hand, research on trust and growth places trust and social capital at the center of economic relations, as facilitators of economic interactions and growth, which relates to the idea developed here. This literature reveals that trust is associated with higher levels of economic growth, because it reduces transaction costs and favors the emergence of efficient institutions (Knack and Keefer, 1997; Easterly et al., 2006; Tabellini, 2008; Algan and Cahuc, 2014). This literature also seeks to determine the origins of trust and cooperation and has emphasized the importance of community characteristics, such as inequality (Alesina and La Ferrara, 2002) or ethnic fractionalization (Algan et al., 2016). Yet, local living conditions, captured by the structure and the quality of the surrounding environment are not considered as potential factors facilitating trust and cooperation. My current research intends to fill this gap by investigating how the "physical" urban environment can increase and shape social encounter, thus contributing (positively or negatively) to social cohesion. In addition, the scope of my research is slightly larger: social cohesion encompasses other dimensions than trust and cooperation. Overall, by exploring the link between urban amenities and economic outcomes through the lens of social cohesion, my research proposes a novel approach placing it at the edge of these two separate streams of research in economics.

Besides, my research relates to a vast literature in other social sciences, notably in sociology, social psychology, and behavioral science. The idea of social cohesion dating back to the dawn of modern sociology in the late 19th century, it would be bold to overlook this body of research. To a large extent, this literature has sought to understand the determinants of social cohesion. In the context of my current research, it is particularly relevant to refer to urban sociology or architecture research, which has put forward the relationship between urban space organization and social interaction (Hanson and Hillier, 1987; Legeby et al., 2015).

My research cannot ignore this rich literature and actually builds on it. Not only do I rely on existing work for defining social cohesion and come out with relevant indicators and measures, but I also take stock of its results in order to formulate assumptions on how urban amenities could affect social cohesion. Nonetheless, I argue that my research also contributes to this literature in at least two important dimensions. First, in addition to understanding the role of urban infrastructure on social cohesion, my research also aims at exploring the effect of social cohesion on economic outcomes. To the best of my knowledge, the existing literature only views economic conditions as a determinant of social cohesion, but not as a possible outcome influenced by tighter and more trustful communities. Second, one of the main challenges I face is to estimate unbiased causal effects. In particular, most of the relationships I study are likely to be biased by endogenous location selection of individuals. My attempt to rely on different empirical strategies (depending on the question at hand, the institutional context and the data available) in order to estimate unbiased causal relationships represents an important dimension of my work.

### **1.1 What is (or not) social cohesion? How can we measure it?**

Before turning to a brief summary of my research, it is in order to define social cohesion. The concept of "social cohesion", that can be traced back to Emile Durkheim's "collective consciousness", has been extensively studied in the academic sociology and social psychology literatures (see [Beauvais et al., 2002](#), for a review). Economics research on social cohesion is more scarce and addresses different issues than the other social sciences do: while sociologists mostly seek to understand whether and how "social bonds" could be maintained in an increasingly industrial and globalized world, economists rather aim at understanding how social cohesion influence economic outcomes. For instance, [Easterly et al. \(2006\)](#) show that social cohesion, defined as "social and economic divisions" positively influences economic growth through better institutions. In addition, over the past three decades, this notion has been heavily relied upon and referred to in the policy discourse, appearing explicitly on many national and international organizations' agendas.<sup>1</sup>

Yet, the definition of social cohesion varies widely, not only between academic and policy papers, but also across different research fields and even within disciplines. In particular, as pointed out by [Chan et al. \(2006\)](#), most of the proposed definitions both refer to the process towards achieving a certain idea of a so-

---

<sup>1</sup>For instance, as reviewed by [Jenson \(2010\)](#), social cohesion is viewed as a main policy goal by the European Union, as emphasized in the Maastricht Treaty in 1992 and in the subsequent Lisbon agenda in 2000; and was called for by the OECD in 1996 in the conclusion of a global conference aiming at defining a social policy agenda beyond 2000.

ciety, and the actual state of this society, thus confusing conditions for and consequences of social cohesion with its actual content. After critically reviewing the definitions proposed in the academic and policy literatures, [Chan et al. \(2006\)](#) present three key elements that should be reflected in a comprehensive definition of social cohesion. First, social cohesion represents the extent to which a society is "sticking together", which implies that (i) its members can trust, help and cooperate with each other, (ii) they share a common identity or feel that they belong to the same community, and (iii) they actually behave accordingly (i.e. these values and norms translate into observable behavior). Second, social cohesion is a state of affairs, and not a dynamic process, so that indicators of social cohesion should reflect attributes rather than conditions or consequences. Thus, although indicators such as unemployment rate or inequality are frequently used in the policy discourse as measures of social cohesion, they are not entirely satisfactory. Indeed, they reflect conditions for, means toward or consequences of social cohesion, but they do not directly reflect the capacity of a society to stick together. Third, members of a society are not only individuals, but also the various organizations and governments (whether local, regional or national) making up the society. Therefore, the notion of trust mentioned previously concerns both horizontal relationships between individuals, and vertical relationships between individuals and organizations. Having these elements in mind, [Chan et al. \(2006\)](#) eventually come up with the following definition of social cohesion, on which I am relying in my various projects: "Social cohesion is a state of affairs concerning both the vertical and the horizontal interactions among members of society as characterized by a set of attitudes and norms that includes trust, a sense of belonging and the willingness to participate and help, as well as their behavioral manifestations." I try to stick as closely as possible to this definition in my research. It is also worth noting that although social capital is a closely related concept, it might not necessarily reflect social cohesion. In a highly ethnically segregated society, for example, individuals may maintain large amounts of networks with members of the same ethnic group even though there may be no relationship across ethnic groups. As [Putnam \(2000\)](#) puts it, there can be a lot of exclusive "bonding capital" (within groups) without much inclusive "bridging capital" (across different groups).

Following this discussion, I now propose a set of relevant social cohesion indicators:

- Local crime, delinquency and insecurity can be used as indicators of horizontal social cohesion, as they reflect low levels of trust and social norms.
- Residential mobility and homeownership rate can be good proxy of local horizontal social cohesion as they are highly related to the sense of belonging to a community (here, the neighborhood).

- Participation to elections (whether local, regional or national) primarily indicate the extent of vertical cohesion. The outcome of the vote, e.g. the share of votes obtained by populist parties may rather reflect the extent of horizontal cohesion.
- Registration on electoral lists can be an alternative measure of social cohesion, reflecting both the sense of belonging to the community and trust towards institutions.
- The number of local associations engaged in solidarity activities, participation and contribution to such associations can be good proxies for horizontal social cohesion, reflecting the sense of belonging to a community.
- Health indicators such as the share of individuals on anti-depressants or the suicide rate can reflect the extent of social integration and of social control as well as the strength of social norms, thus reflecting local horizontal social cohesion.
- Intensity and quality of the relationships with neighbors also reflect local horizontal social cohesion.

So far, I have relied on several of the above indicators for social cohesion, but not all of them. This leaves some avenue for future research. For each paper, I will mention the exact source from which I computed the indicator under study.

## **1.2 Research overview**

My research is structured around two main axes: 1/ Understanding the relationship between social cohesion (particularly local social interactions), the local economic context and the economic situation of urban residents; 2/ Understanding the role played by the urban environment on economic outcomes and social cohesion within cities.

### **1.2.1 Local social cohesion and economic conditions**

The first axis of my research aims to understand the links between local social cohesion and local economic conditions, and to identify the direction of causality between these two components, which can go in both directions. Previous research in economics suggests a positive effect of trust ([Algan and Cahuc, 2010](#)) and social cohesion more generally ([Easterly et al., 2006](#)) on economic outcomes, and in particular on growth, at the aggregate (generally cross-country) level. According to this literature, more integrated or trustful nations allow for the emergence of better-functioning institutions, leading to better economic outcomes. We can therefore assume that similar mechanism is at play at a more local level: socially cohesive communities,

characterized by higher levels of trust, cooperation, and stronger common norms should be more likely to develop or maintain efficient local institutions, leading to better economic outcomes. For instance, local social connections have been shown to be particularly relevant in explaining labor market outcomes (see for instance [Ioannides and Datcher Loury \(2004\)](#) for a review and [Solignac and Tô \(2018\)](#) for a recent study in France). I have been investigating this kind of relationship in several papers that I briefly present here and to which section 2.1 will be devoted.

In particular, my research focuses to a large extent on social interactions as a measure of social cohesion, and explores how it affects various social and economic outcomes, at the local level. The importance of social interactions in individuals' lives has a long standing tradition in sociology (see for instance [Granovetter, 1973](#); [Montgomery, 1994](#)) and has been extensively investigated in the economics literature as well. Whether they occur in the public space (classroom, firm, public transportation, neighborhood...) or in the private sphere (e.g. between friends or family members) social interactions are known to influence a variety of social and economic outcomes: education ([Calvó-Armengol et al., 2009](#); [Giorgi et al., 2010](#); [Lin, 2010](#); [Boucher et al., 2014](#); [Patacchini et al., 2017](#); [Altonji and Mansfield, 2018](#); [Mendolia et al., 2018](#)), crime ([Glaeser et al., 1996](#); [Calvó-Armengol and Zenou, 2004](#); [Damm and Dustmann, 2014](#)), technology adoption ([Conley and Udry, 2010](#)), consumption ([De Giorgi et al., 2020](#)), smoking ([Fletcher, 2010](#)), and labour market outcomes (see [Bayer et al., 2008a](#); [Hellerstein et al., 2011](#); [Dustmann et al., 2016](#), for general evidence of the existence of informal hiring networks) to name only a few.<sup>2</sup> The common idea in this wide literature is that non-market interactions are important in determining social and economic outcomes by influencing individuals' beliefs, knowledge or social norms more generally.

In line with this literature, an important part of my research has been devoted to studying the role of social relationships in the economic sphere, and to uncovering the underlying mechanisms. This work will be presented in section 2.1. My first papers study the role of local social interactions in the specific context of diversity in terms of cultural, legal or ethnic origins, having in mind the idea that neighborhood's composition is likely to influence interactions among neighbors.

In my paper entitled "**The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation**", written with Yann Algan (Sciences Po) and David Laitin

---

<sup>2</sup>See [Topa and Zenou \(2015\)](#) for a review on neighborhood and network effects.

(Stanford University), and published in the *Journal of Political Economy*, we provide some evidence of the role of social relationships on the quality of public spaces and housing in French public housing blocks, in a context of diversity of origins (Algan et al., 2016).<sup>3</sup> Our results reveal that diversity has a negative effect on the quality of local public goods, which is explained not only by acts of vandalism linked to a lack of peer pressure, but also by ineffective management of the social housing stock due to the failure of collective action in the most diverse neighborhoods. By contrast, we also show that diversity has no negative impact on public safety. This paper will be presented in section 2.1.1.

A second part of my research on the economic effects of social interactions alternatively focuses on the role of interpersonal relations on employment. This body of research will be presented in section 2.1.2. There is considerable evidence that information transmission plays a key role on the labor market. Early empirical studies based on survey data conducted over various time periods and on diverse countries agree that relying on friends and family is a very popular job search method and that on average half of jobs are found through social networks (Corcoran et al., 1980; Granovetter, 1995; Holzer, 1988; Wahba and Zenou, 2005). Theoretically as well, Calvó-Armengol and Jackson (2004) show that employment probability increases both with the number of links an agent has, and with the employment rate in the individual's network. Relying on administrative or matched employer-employee data, subsequent papers have contributed to indirectly detect informal referral networks. In their seminal work, Bayer et al. (2008a) show that neighbors living in the same census block are much more likely to work at the same location than residents of nearby blocks, a tendency that is confirmed by Hellerstein et al. (2011). On top of confirming previous findings on the importance of informal hiring networks, these papers suggest that individuals share employment opportunities with their immediate neighbors and that overall, referral networks have an important spatial dimension. Focusing on displaced workers during the Great Recession, recent work by Hellerstein et al. (2019) confirms the key role of local acquaintances in job finding.<sup>4</sup>

Following my first paper presented above, I first aimed at understanding the relationship between social interactions and employment in a context of diversity of origins. In my paper entitled "**Diversity and Employment Prospects: Neighbors Matter!**", joint with Clément Malgouyres (Institut des Politiques

---

<sup>3</sup>This paper was started during my PhD studies but subsequently went through important revisions during my post-doctoral period.

<sup>4</sup>A parallel strand of the literature has investigated the role of informal hiring networks on the quality of the employment match. For instance, Schmutte (2015) finds that workers whose neighbors work in high-paying firms are more likely to find jobs with higher wages, and Dustmann et al. (2016) provide evidence that workers earn higher wages when they obtained their job through a referral.

Publiques and Paris School of Economics) and published in the *Journal of Human Resources*, we estimate the effect of a community's level of diversity on the employment prospects of its members (Hémet and Malgouyres, 2018). Our results show that, at the neighborhood level, diversity has a negative effect on individual employment. Our interpretation is that lower communication across groups of different origins hinders the transmission of job information, thus reducing the employment prospects of residents in more diverse neighbourhoods. By contrast, we do not find any significant effect of diversity on employment at the commuting zone level. Then, in my working paper entitled "**Ethnic Networks and the Informal Labor Market**", for which revisions were requested at the *Journal of Public Economic Theory*, I develop a dynamic theoretical model to study the role of ethnic networks on the propensity to hold an undeclared job (Hémet, 2017). The model assumes that networks feature homophily (along ethnic lines), and reveals that minority workers turn to informal employment because they lack of formal job opportunities due to their smaller network. Finally, in a recent working paper entitled "**The Strength and Use of Local Referral Networks**", joint with Clément Malgouyres, we propose an empirical analysis of the general role of neighbors as job referrals, going beyond the context of diversity of origins (Hémet and Malgouyres, 2019). We estimate the propensity of immediate neighbours to work in the same establishment compared to neighbours in adjacent neighbourhoods, in the spirit of Bayer et al. (2008a). Our preliminary analysis seems to confirm the presence of local networks of job recommendations. The next part of this project will explore to what extent more horizontal social cohesion (measured by participation in local associative and civic life, for example) reinforces these network effects.

Eventually, as a follow-up of the previous papers on social interactions and economic outcomes, I am starting to work on two projects focusing on educational outcomes. These two projects at their onset will be briefly exposed in section 2.1.3. In one project conducted with Manon Garrouste (University of Lille), we aim at understanding the influence of local peers on orientation decisions. In another project joint with Arnaud Chevalier (Royal Holloway, University of London) and Matthijs Osterveen (University of Lisbon), we propose to investigate experimentally the mechanisms underlying the relationship between classroom diversity and academic achievement.

Regarding the reverse relationship between economic conditions and social cohesion, existing literature in various social sciences reveals that economic inequality is one of the strongest predictors of cross-country

differences in social trust (see for instance Bjørnskov 2007, Rothstein and Uslaner 2005). Following my first body of research presented above, I seek to understand this inverse causal relationship, *i.e.* the influence of local economic conditions on social cohesion. This ongoing research will be exposed in section 2.2.

In a working paper entitled "**Unemployment and Crime Victimization: A Local Approach**", presented in section 2.2.1, I estimate the effect of local unemployment on the probability of being a victim of crime in one's neighbourhood, where crime can be viewed as an indicator of social cohesion as explained above (Hémet, 2018). Consistent with the hypothesis proposed by (Becker, 1968), according to which the decision to commit a crime results from a rational economic cost-benefit analysis, I find that local unemployment rate increases crime victimization. In particular, unemployment in one's own neighborhood relates to vandalism, while unemployment in adjacent areas rather relates to profit-driven crimes.

On another note, research in economics has shown that economic shocks, such as trade shocks induced by the rise of Chinese exporting industries, are associated with a decline in local spending on public goods and services (see for instance Feler and Senses (2017) for the US). Following these results I will briefly expose a future project joint with Clément Malgouyres, in section 2.2.2. Our objective is to explore the impact of local economic shocks (e.g. related to firms' exposure to international trade opening) on voting behavior, another indicator of social cohesion. In particular, we plan to investigate whether local public spending (e.g. investment in local amenities and public goods) can contribute to mitigate the effect of economic shocks on social cohesion.

Finally, a related though slightly different question lies in the role of the institutional setting (here, the features of a justice system) in fostering social cohesion. In a recent working paper, entitled "**Custodial versus Non-Custodial Sentencing: Evidence from an Anticipated Reform**", joint with Bastien Michel (Paris School of Economics), we take advantage of a reform of the Danish legislation on drink-driving crimes to study the impact of custodial (e.g. incarceration) versus non-custodial sentences (e.g. community services) on offenders' life trajectories (Michel and Hémet, 2019). We find that, compared to custodial sanctions, non-custodial ones reduce the number of subsequent crimes and improve labor market outcomes. This piece of research will be presented in section 2.3.

### 1.2.2 Urban environment and socio-economic outcomes

The second axis of my research aims at understanding how the city, through its infrastructures, its amenities and its form, influences socio-economic interactions and, by extension, economic conditions. We can think of several ways through which such a relationship could exist. First, some types of equipment or infrastructure may play a role by facilitating social interactions, within or across groups (generally defined along social or ethnic lines). For instance, public facilities such as green spaces (parks) and local consumption amenities such as bars and restaurants are a vector of increased social encounters: all else equal (and in particular for a given local social or ethnic composition), individuals living in areas containing such facilities can meet more easily. Such relationship between urban space organization more generally and social interaction has been illustrated in urban sociology or architecture research, relying on case studies (see for instance the early work by [Hanson and Hillier \(1987\)](#) or the recent study by [Legeby et al. \(2015\)](#) in Gothenburg). Similarly, transportation infrastructure, whether highways or public transportation networks can not only increase the frequency of social encounters (*e.g.* through reduced transportation cost and time), but also enlarge the scope of interactions. Indeed, transportation networks can connect socially deprived areas to richer ones, thus enabling spatially and socially distant individuals to interact, and may improve social cohesion at a more aggregate level (*i.e.* beyond the boundaries of the local neighborhood).

Yet, even if the presence of some amenities favor social interaction, it is not clear how this would affect measures of social cohesion. More frequent interaction could improve local social cohesion positively through increased trust or cooperation for instance (due to, *e.g.* repeated interactions). On the other hand, if these interactions remain superficial, this could lead to a reinforcement of negative stereotypes about individuals belonging to different groups ([Uslaner, 2012](#)), thus impeding social cohesion. Similarly, getting socially distant communities closer through better transportation infrastructure can emphasize this distance and put more strain on deprived individuals, leading them to criminal behavior ([Merton, 1938](#)). By contrast, renewal policies aiming at improving the deteriorated housing stock in deprived neighborhood can attract richer households, therefore increasing the social mixing (see [Guyon, 2017](#), for a report on France) on a more permanent basis, which would rather play towards more integrated and cohesive societies. On another note, the access to and quality of local public goods services (ranging from clean streets to access to good schools and health services) can affect individuals' trust in the government or reinforce their sense of belonging to a common community. More generally, if amenities (consumption, transportation, parks, access

to public services) increase residential stability (Rappaport, 2009), they may contribute to create a common identity and sense of belonging to the same community.

My research addresses various aspects of this broad question. In a first set of papers joint with Miquel-Àngel Garcia-López (Universitat Autònoma de Barcelona and Institut d'Economia de Barcelona - IEB) and Elisabet Viladecans-Marsal (University of Barcelona and IEB), we estimate the impact of transport infrastructure on the location and local growth of population and employment, and on the shape of the city. To do so, we focus on the Paris metropolitan area (the *Ile de France* region), which has undergone major road and rail network changes over the past fifty years. Our results highlight the predominant role of the Regional Express Rail (RER) in the urban dynamics observed over the period 1968-2010. In a first paper entitled "**How does Transportation Shape Intrametropolitan Growth? An Answer from the Regional Express Rail**", published in the *Journal of Regional Science*, we essentially show that the proximity of a RER station increases employment, population density and growth rate of these two variables (Garcia-López et al., 2017a). Moreover, in a companion paper entitled "**Next Train to the Polycentric City: The Effect of Railroads on Subcenter Formation**" and published in *Regional Science and Urban Economics*, we show that the RER has fostered the emergence of suburban employment centers, thus contributing to the decentralization process in the Paris metropolitan area and to the emergence of a polycentric city (Garcia-López et al., 2017b). Both papers are presented in section 3.1.

After considering the role of transportation infrastructure on the dynamics and structure of employment at the metropolitan level, my research currently aims at understanding how urban amenities and infrastructures (what I call the "urban environment") influence social cohesion. In an ongoing project joint with Nina Guyon (National University of Singapore) and Arnaud Philippe (University of Bristol), our purpose is to explore whether urban renewal programs, whereby urban amenities and housing quality are improved, have an effect on social cohesion. To this aim, we evaluate the impact of the French urban renewal program that was conducted between 2004 and 2015 on crime. In the medium run, we also plan to look at the effect of this program on election turnout and share of votes obtained by the main extreme-right party, as alternative indicators of social cohesion. Finally, as a natural follow-up of the various projects presented in this section, I plan to devote a future project to exploring the impact of transportation infrastructure on crime. Both

projects are presented in section 3.2.

The remainder of this dissertation is organized as follows. Sections 2 and 3 are devoted to present my research pertaining to the first and second axis respectively. Eventually, I will describe my other research-related activities in section 4.

## **2 Local social cohesion and economic conditions**

In a first body of research, I investigate the relationship between social cohesion interactions and economic conditions. This relationship can be considered in a dual way: improving social cohesion may affect economic outcomes, and, conversely, economic conditions may influence social cohesion. My research looks into both directions, as is detailed in section 2.1 for the former – with a specific focus on social interactions as an indicator of social cohesion – and 2.2 for the latter. A recent piece of research presented in section 2.3 gives a hint at both sides of the question at once, although in an indirect way.

### **2.1 On the influence of social interactions on social and economic outcomes**

An important part of my research is devoted to studying the role of social relationships in the economic sphere, and to uncovering the underlying mechanisms. So far, I have been focusing on housing conditions and on labor market outcomes. Part of my current (ongoing) research is directed towards educational outcomes. These three blocks will be described in section 2.1.1, 2.1.2 and 2.1.3 respectively. The common denominator of my research lies in the fact that I study very local interactions (*e.g.* neighborhood level). In some instances, this local standpoint enables me to contribute to branches of the literature that have mostly investigated the question from a more aggregate perspective. In all papers, this has also the advantage of allowing me to provide some insights on the mechanisms. Yet, working at very low spatial levels systematically raises the issue of individuals' sorting (*e.g.* residential sorting), thus jeopardizing the causal interpretation of the empirical results. In order to bypass this issue and estimate unbiased effects, I propose various identification strategies in my different papers.

### 2.1.1 Interactions and housing conditions in a context of diversity of origins

A large literature shows a negative relationship, though not always robust, between ethnic diversity and several other outcomes: quality of public goods (Alesina et al., 1999; Alesina and La Ferrara, 2000; Miguel, 2004; Miguel and Gugerty, 2005), welfare spending (Luttmer, 2001), civil conflict and trust (Fearon and Laitin, 2000; Putnam, 2007; Alesina and La Ferrara, 2002) and economic growth (Alesina et al., 1999). This literature focuses on rather aggregate outcomes, and considers the problem of diversity in terms of voting behavior on aggregate outcomes such as public goods provision at the country or county level. Yet, these papers do not take a stand on how diversity affects social relationships and well-being at the neighborhood level.

In a paper entitled "**The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation**", written with Yann Algan (Sciences Po) and David Laitin (Stanford University) and published in the *Journal of Political Economy* (Algan et al., 2016), we contribute to the previous literature by analyzing how diversity within a small community affects individual well-being and satisfaction with housing conditions, through the lens of social relationships among neighbors. To this aim, we rely on very detailed data on housing conditions from the French Housing Survey, where we know respondents' precise block of residence (*ilot* in French), and that we can match to representative block level census data. In particular, we compute the level of diversity in terms of nationality at birth. The variety of outcomes available from the housing survey enable us to estimate various diversity effects and to get a sense of the underlying mechanisms (as will be explained below), contributing further to the literature.

The likely endogenous residential sorting of individuals on ethnic grounds implies that any naive estimate of the effect of diversity on housing conditions might be biased. We address this issue by relying on the population of social housing tenants, who are allocated to public housing blocks irrespective of their ethnic origin or their preference for diversity. This implies that we can take the degree of diversity in any one public housing block as exogenous. We provide several arguments supporting this assumption. In particular, we extensively document the actual process of allocation of households within the public housing sector. We show that legal rules prohibit housing allocation based on ethnic backgrounds and that in practice, the characteristics of the public housing sector make it very complicated to bypass the law. We also conduct a variety of formal statistical tests to verify the absence of self-sorting on ethnic characteristics.

Focusing on the population of public housing tenants, for whom diversity can be considered exogenous,

we are able to estimate the effect of diversity on three main dimensions of housing quality that lead us to derive alternative interpretations and get some insights on the mechanisms. We first find a positive effect of diversity on outcomes capturing voluntary degradation in the common areas, such as graffiti or broken mailboxes, for which tenants can be held responsible. Then, we also estimate a positive effect on outcomes signaling poor quality of housing, such as poor elevator maintenance or heating-related problems, that can neither be caused nor solved by the tenants, but for which public housing offices are responsible. Finally, we do not find any significant effect of diversity on outcomes related to public safety, that refer to interpersonal aggression and criminality in the neighborhood.

Considering that a third of public housing tenants report having no relationship at all with their neighbors, the most plausible interpretation that encompasses findings for all three broad types of outcomes, is that diversity generates social anomie, *i.e.* the absence of common rules and social norms. As a consequence of anomie, there are (a) weaker other-regarding preferences and a lack of credible threats of social sanctions, hence more neglect and vandalism, (b) a failure to generate collective action to pressure the public housing offices into improving housing quality, and (c) fewer opportunities for violent interpersonal confrontation.

### **2.1.2 Social interactions and employment - *Working and published papers***

A second part of my research focuses on the role of social interactions on employment. Following the paper presented above, I first addressed this question in the specific context of diversity of origins. Ethnic networks matter substantially for minorities employment prospects, as revealed by research in economics (Battu et al., 2011) and in sociology (Light and Rosenstein, 1995; Light, 2005). This could be explained by particularly strong information transmission within ethnic groups (Fearon and Laitin, 1996; Bertrand et al., 2000). As a consequence, ethnically or culturally diverse neighborhoods may be characterized by lower information transmission, which could in turn affect residents' employment prospects.

In a published paper entitled "**Diversity and Employment Prospects: Neighbors Matter!**", joint with Clément Malgouyres (IPP, PSE) and published in the *Journal of Human Resources*, (Hémet and Malgouyres, 2018), we empirically investigate the effect of living in a diverse environment on employment probability. In particular, we address this question through the lens of local social interactions. Existing literature does not provide a clear picture of the direction of this effect, as several mechanisms can be at play. At the micro

level, if members of different ethnic or cultural groups barely interact with each other, then diversity can act as a barrier to communication and in particular to job information transmission. Given the importance of personal networks in the job search process discussed above, diversity could therefore reduce the chance of finding a job. On the other hand, if communication across groups is not an issue, then mixing people conveying non-redundant pieces of information (*e.g.* due to different backgrounds) can certainly improve employment prospects. At a more aggregate level, diversity can affect employment probability through its impact on productivity, leading to ambiguous effects as well. On the bright side, diversity can be beneficial to productivity due to complementarity in workers' skills (Lazear, 1999; Alesina and La Ferrara, 2005, see). On the downside, heterogeneity can hinder productivity by preventing social capital formation (Coleman, 1988).

In this paper, we use detailed data from the French labor force survey to measure diversity using several definitions of origins and at various geographic levels, so as to understand as precisely as possible the mechanisms lying behind the diversity-employment relationship. We are thus able to disentangle the neighborhood (local) and local labor market (aggregate) effects of diversity, which is a key contribution to the literature. Specifically, for each spatial level, we rely on three alternative definitions of origins to measure diversity: nationality, birth country, and parents' origins.<sup>5</sup> The measure based on parents' origins encompasses first- and second-generation immigrants and is therefore more closely related to the standard ethnic classification of individuals. By contrast, defining kinship according to nationality introduces the notion of immigrants' integration through naturalization. This distinction allows us to draw conclusions on whether the cultural or the ethnic dimension of diversity prevails.

Our study is not exempt from the usual concern that individuals may have a preference for living close to their co-ethnics, thus biasing any measure of the effect of diversity. Another issue is that of reverse causality that can arise if immigrants decide to settle in more economically dynamic areas. We propose several identification strategies to bypass these concerns, depending on the level of aggregation considered. The endogeneity of local labor market diversity is handled through a traditional instrumental variable approach, where two different instruments are proposed. First, following the "shift-share" approach proposed by Card (2001), we construct the predicted level of diversity in each employment zone based on the distribution of

---

<sup>5</sup>By contrast, the existing literature studying the effects of diversity mostly relies on ethno-linguistic and religious groups.

each origin group across employment zones in 1968 and the current number of individuals from each origin in France at the time of the study (2007-2013). An alternative and more innovative instrument is the level of diversity within the public housing tenants of the employment zone. It builds on my work presented above in section 2.1.1 (Algan et al., 2016), which shows that the allocation of households across public housing units in France does not take their origins or their preference for diversity into account, so that diversity can be considered as exogenous in the public housing sector. Building on this result, we argue that the general level of diversity in a local labor market partially determined by that in the public housing sector, both mechanically (social housing tenants are included in the general population of the area, so that their composition affects the composition of the overall population) and through a magnet effect on immigrants living in the private sector (residents in the private market, who are less constrained on their location choice, might be attracted to an area where some of their co-ethnics were located through the public housing allocation process). We therefore use public housing diversity to instrument the general level of diversity at the local labor market level. Indeed, because the public housing population is part of the total population, fractionalization based only on this sub-population is mechanically proportional to fractionalization based on the population as a whole. Also, it is reasonable to think that individuals living in the private housing market, and who are therefore less constrained upon their location choice, are influenced by the composition of the public housing population. Otherwise stated, people from a given group might be attracted by an area where some of their co-ethnics were located through the public housing allocation process, so that diversity in the area is likely to reflect diversity in the area's public housing sector.

To deal with the endogeneity of neighborhood diversity, we follow Bayer et al. (2008a) and assume that although households are able to select the precise area in which they want to live, they are, however, unable to pinpoint an exact neighborhood within this given area. Therefore, after controlling for sorting in a restricted area encompassing local neighborhoods, the assignment of individuals to a specific neighborhood within the selected area is considered as essentially random and provides a credibly exogenous source of variation to identify the effect of diversity. In the context of this paper, this identification strategy hinges on the very specific sampling scheme of the French labor force survey (LFS) that makes it extremely valuable for studying neighborhood effects. The baseline sample is made of about 2,500 areas partitioning the country, called sectors, which are split into six contiguous areas of 20 homes on average, corresponding to what we refer to as "local neighborhoods". The LFS being a rotating panel, the first neighborhood of a sector is

exhaustively surveyed over six quarters, before being replaced by the second one for the next six quarters. This process goes on until all six neighborhoods of the sector are surveyed. This particular sampling strategy implies that we observe all individuals living in these local neighborhoods, so that we can perfectly characterize their socio-economic and demographic environment. Interestingly, it also implies that we are able to observe individuals living in contiguous local neighborhoods within the same sector, a feature that is key to our identification of neighborhood diversity effects.<sup>6</sup>

Naïve OLS estimates reveal that employment probability is negatively correlated with neighborhood diversity, but positively correlated with employment zone diversity, suggesting a negative effect through networks and a positive one through productivity. Once we account for the potential endogeneity of location across local labor markets, we do not find robust evidence of a causal impact of employment zone diversity on individuals' employment status: none of the IV estimates are statistically significant and they all have large standard errors despite strong first stages. The imprecision of the estimate hence precludes any firm conclusion regarding the absence of a causal effect of diversity at the aggregate level. However, additional results obtained using the specific subsample of public housing tenants, for whom diversity can be considered as exogenous at the block level, point toward a null effect of aggregate diversity and tend to support the intuition that the positive aggregate relationship suggested by OLS estimates was actually driven by selection.

Now turning to more local estimates, we find that, after controlling for residential sorting through the inclusion of *sector* fixed effects, the effect of neighborhood diversity remains significantly negative, though slightly attenuated. This indicates that sorting resulted in more economically deprived individuals living in more diverse neighborhoods. Interestingly, we also observe that diversity based on nationality has a larger impact than diversity based on birth country, which is itself more relevant than diversity based on parents' origins. We argue that parents' origins more likely reflect ethnicity than nationality, the latter being rather associated to common values and language. An interpretation of these results is therefore that neighborhood diversity lowers the probability of employment because of cultural differences, most likely including language differences, rather than ethnic differences. This reinforces the intuition that diversity affects job finding by limiting job information transmission between neighbors.

---

<sup>6</sup>In addition to explaining this particular sampling scheme in great details, the paper also proposes statistical tests of the identifying assumption.

I subsequently investigate the related question of the role of networks in finding regular versus undeclared jobs, in the context of ethnic diversity. Given the challenge represented by obtaining micro-data on the informal economy, this question is addressed from a theoretical perspective in a working paper entitled "**Ethnic Networks and the Informal Labor Market**" (Hémet, 2017). In a setting close to [Calvó-Armengol et al. \(2007\)](#) who study the role of interpersonal relationships on the decision to commit a crime, I develop a dynamic model of the informal labor market. Specifically, agents belonging to two different ethnic groups (*e.g.* minority and majority workers) randomly meet by pair at each period, and exchange information on two competing activities: declared and undeclared employment. Formal workers pay taxes, while informal workers evade them. Workers from each group are homogeneous in terms of skills, and can be thought of as low-skilled workers without loss of generality.

In this model, networks play a central role, as information about both formal and informal job opportunities can only be obtained through word of mouth communication, in line with the empirical evidence mentioned above about the essential role of networks in job finding. In addition, social interactions are characterized by homophily, that is, individuals tend to form links and gather with people similar to themselves, and in particular along ethnic lines (see [McPherson et al. \(2001\)](#) for a review of the literature on homophily, and [Currarini et al. \(2009\)](#) and [Bramoullé et al. \(2012\)](#) for models of network formation featuring homophily). As a consequence of this assumption, information about job offers circulates imperfectly across groups.

I analyze the flows of individuals between the various occupations (namely unemployment, formal employment and informal employment) and characterize all the steady-states of this dynamic economy. For this purpose, I solve for the endogenous individual decisions to accept or reject formal and informal job offers. I then derive the conditions under which these steady-states arise in equilibrium. Interestingly, when the net formal wage is larger than the net informal wage, all equilibria involving participation in the informal economy arise for low enough values of the unemployment benefit. This reveals that when the unemployment benefit is too low, individuals are more likely to immediately accept an informal job offer when they receive one rather than to decline and wait for a formal job opportunity.

In particular, an equilibrium in which both groups are formally employed but one also engages in the informal sector arises, and I show that this group is the minority group. This equilibrium is in line with the intuition that the informal sector may be used as an economic safety net when formal employment prospects

are low. Indeed, the minority group faces such a low rate of formal job offers arrival (due to the size of the network) that its members are better off immediately accepting an informal job offer than waiting longer for a formal one. On the contrary, unemployed members of the majority group, whose network is more developed and hence who are more likely to receive formal job offers, can afford declining unattractive informal offers.

The model also shows that policies reducing homophily, that is enhancing inter-groups relationships and improving the integration of minority workers tend to favor formal over informal employment. More importantly, such policies turn out to be neutral in terms of unemployment, as they directly amplify the flow toward the formal sector. By contrast, policies directly combating undeclared jobs through more frequent audits seem less efficient: they do limit the size of the informal economy, but at the expense of a larger unemployment rate. Indeed, such policies boost the flow out of informality, but have no direct impact on the flow into formal employment.

I eventually investigate the relationship between social interactions and employment, going beyond the specific context of diversity of origins. In a recent and preliminary working paper entitled "**The Strength and Use of Local Referral Networks**", joint with Clément Malgouyres ([Hémet and Malgouyres, 2019](#)), we present new evidence individuals rely on their employed local neighbors to find jobs (what we call "local referral networks") in the French context. Using the French Labor Force Survey, which contains exhaustive information on close neighbors as well as employer administrative identifiers, we measure the strength of informal hiring network as the differential propensity of individuals living in the same block to work in the same firm in comparison with residents of a contiguous block, in the spirit of [Bayer et al. \(2008b\)](#). More precisely, we estimate that being close neighbors increases the probability of being coworkers relatively to more distant neighbors by 0.35 % on average, which represent about a third of the baseline probability that neighbors work in the same firm.<sup>7</sup>

In order to bypass the problem of endogenous residential sorting, we assume that although households are able to select the precise area in which they want to live, they do not, however, chose upon the exact neighborhood where they end up living within this given area. This identifying assumption is similar to one that is used in [Hémet and Malgouyres \(2018\)](#) and that was presented at the beginning of this section 2.1.2. This assumption implies that, after controlling for sorting in a restricted area encompassing local neighbor-

---

<sup>7</sup>Interestingly, this result is very similar to that obtained by [Bayer et al. \(2008a\)](#) using US data.

hoods, the assignment of individuals to a specific neighborhood within the selected area is considered as essentially random and provides a credibly exogenous source of variation to identify neighborhood effects. Indeed, we show that while the raw data exhibits a large amount of sorting on observables, there is virtually no correlation between individual and neighborhood-level predetermined variables (such as educational achievement or being foreigner) left once we condition on fixed effects for the larger area that households chose. The absence of within-area sorting on observables supports the validity of our identifying assumption.

In order to deal with the issue of reverse causality, i.e. the possibility that people might be meeting at work and referring each other on the housing market, we first estimate our empirical model on a sub-sample of public housing tenants. Given the shortage of public housing as well as the allocation mechanism of public housing units explained in section 2.1.1 above (Algan et al., 2016), the risk of reverse causality is non-existent in this setting. Second, we exploit the short-longitudinal dimension of our data to measure the role of local referral network on labor market transitions. Looking at workers at the time they enter a new firm rules out the problem of reverse causality albeit at the cost of a substantial reduction in sample size.

The main contribution of this work lies in our ability to look at job transitions and to analyze in which type of transition (job-to-job versus out of non-employment) the use of local referral is the most intensive. We find that local referral network seems to be more relevant for unemployment to job transitions than for job-to-job transitions. By contrast, studies based on administrative data usually focus on job-to-job transitions due to the poor quality of data for people outside of employment (Schmutte, 2015, *e.g.*). Our finding thus suggests they might be missing an important role of local labor market referrals.

It is also worth noting that most evidence on the existence of neighborhood-based referral network is based on US data. Evidence for continental Europe and particularly France is much more scarce and we contribute to fill this gap. A recent paper by Solignac and Tô (2018) focus on young graduates in France and show that neighborhood employment rate positively affects their employment probability. While their work finds evidence of social endogenous effect on labor market entry, our work suggests that this kind of effect could be driven by the existence of direct referrals. Now that we have confirmed the presence of local referral networks through this preliminary analysis, our next step will be to identify the role of social cohesion in this relationship: relying on alternative indicators of horizontal social cohesion at the local level, we plan to estimate how increasing social cohesion reinforces referral networks.

### 2.1.3 Social interactions and education - *Recent ongoing projects*

Eventually, I am currently working on two new projects linking local social interactions and education, an important dimension that was previously absent from my research.

**Identifying the role of peers in education decisions** In an ongoing research project joint with Manon Garrouste (University of Lille), we aim at understanding whether and to what extent interactions with other students in the same neighborhood affect individual education decisions. While it is established that social reproduction is an important dimension of educational track choice in France, much less is known about the way in which this choice can be influenced by neighborhood peers' orientation decisions.

In a large and growing literature focusing on peer effects in education, several studies use the US Add Health Survey to shed light on the role of school peers on educational attainment – usually measured by tests scores, drop out rates and college enrollment (see [Calvó-Armengol et al., 2009](#); [Lin, 2010](#); [Bifulco et al., 2011](#); [Patacchini et al., 2017](#), among others). Contrasting with these papers which look at influence from *school* peers, [Goux and Maurin \(2007\)](#) rely on French data to explore *neighborhood* peer effects. They show that close neighbors' characteristics influence test scores and grade repetition, but their setting does not allow them to take a stand on orientation decision. Few papers directly look at track choice, and if so, they document the role of school peers' characteristics (e.g. educational attainment and social background) rather than peers' choices ([Jonsson and Mood, 2008](#); [Speckesser and Hedges, 2017](#)). Instead, we want to take a stand on how students' decision making is influenced by peers choices.

This question was already addressed by [Giorgi et al. \(2010\)](#) who show that one is more likely to choose a given college major when many of her peers make the same choice. While they focus on major choice for undergraduate students in a prestigious Italian university, we will rather study orientation decision for students at the end of ninth grade. In France, this is a key timing in the education path, as students may decide to embark upon vocational training rather than academic track. The decision made at the end of ninth grade thus deeply influences future education and labor market outcomes. For that reason, it is crucial to understand how this decision is taken and how it can be impacted by the choices made by other pupils in the neighborhood.<sup>8</sup>

To this aim, we use exhaustive administrative geolocalized individual data provided by the statistical

---

<sup>8</sup>Several closely related papers also look at this type of decision between vocational and academic training, but they focus on the role of information rather than neighborhood peer effects *per se* ([Fricke et al., 2018](#); [Goux et al., 2017](#)).

service of the French ministry of education. This data not only allows us to observe ninth graders' track choice, but also to precisely identify her neighbors and their track choices. This unique setting enables us to take a stand on the question at hand. Crucially, observing students' precise residential location gives us a way to bypass the issue of endogenous location decision that is particularly salient in the context of schooling decisions. Indeed, residential location largely determines which school will be attended, based on its catchment area, so that we expect socially and economically more advantaged families to locate in good quality schools' catchment areas. In turns, this means that location in catchment area is not exogenous to schooling decisions and preferences. We will therefore rely on a fixed effect strategy (previously described in section 2.1.2 and used in Hémet and Malgouyres (2018) for instance), whereby the identifying variation will come from comparing local neighborhoods *within* these highly selected catchment areas.

**Social interactions and educational outcomes in a context of diversity of origins** The second ongoing project, carried out with Arnaud Chevalier (Royal Holloway, University of London) and Matthijs Oosterveen (University of Lisbon), relates to my previous work on ethnic diversity, and seeks to understand the channels through which it influences students' academic achievement. The large body of literature on diversity effects has proposed several mechanisms to explain the observed results. For instance, Miguel and Gugerty (2005) or my previous work presented in section 2.1.1 (Algan et al., 2016) suggest an important role of social sanctions and social norms. On another note, papers such as Alesina and La Ferrara (2005) rather insist on heterogeneous preferences across ethnic groups. Yet, none of these papers actually test the proposed mechanisms, and the literature on this specific topic remains scarce.

This project intends to fill this gap relying on quasi-experimental settings whereby first year students are randomly allocated to tutorial groups in two European universities (Erasmus University, Rotterdam and Royal Holloway, London). Using these settings where group diversity can be considered as exogenous, we will have students play standard behavioral economic games (in the spirit of Fershtman and Gneezy, 2001; Berge et al., 2019) in order to measure their level of interpersonal trust and biases. Importantly, these games will be played repeatedly (three times) throughout their first year, enabling us to investigate how attitudes and biases evolve over time, and whether they evolve differently for students in more diverse tutorial groups. In particular, the first game session will take place at the onset of the academic year, just after tutorial groups have been formed. This will provide us with a baseline measure of trust and biases before students start

bonding with each other.<sup>9</sup> Repeating the games is one of the main innovations that we propose compared to the existing literature where individual responses are usually observed at one point in time. This repeated setting will allow us to observe the evolution of trust, cooperation and ethnic biases over time.

Our project is closely related to two recent papers. [Carrell et al. \(2020\)](#) rely on a quasi-experimental setting whereby U.S. Airforce Academy students are randomly assigned to peer groups, and show that attitudes towards black students is improved after being assigned to groups with more black students. While their paper investigates relationship between ethnic groups that have a long common history in the US, our setting will look at diversity arising as a result of recent immigration in the EU, i.e. in a setting where individuals share less of a common cultural heritage. Our setting is thus closer to that of [Finseraas and Kotsadam \(2017\)](#) who randomly assign Norwegian soldiers into different rooms during their initial training period in the army. They find that natives assigned to dorms with a larger share of individuals from the minority subsequently report better perception of minority individuals' work ethics. Compared to this study, we will make students play experimental games repeatedly over the period, so as to understand precisely the speed at which behavior, preferences and biases towards minorities evolve over time.

## **2.2 On the influence of economic conditions on social cohesion**

### **2.2.1 Spatial economic disparities and crime victimization**

While the works presented above explore how social interactions influence local economic conditions (housing, employment), I have also focused on the reverse causality, i.e. the effect of local economic conditions on social cohesion. In a working paper entitled "**Unemployment and Crime Victimization: A Local Approach**" ([Hémet, 2018](#)), I focus on the role of local economic conditions on a specific indicator of social cohesion, namely the individual probability to be victim of a crime.<sup>10</sup> In particular, I show that unemployment rate has a very local spatial influence that varies across types of crimes.

Seminal papers by [Becker \(1968\)](#) and [Ehrlich \(1973\)](#) hypothesize that the decision to engage into criminal activities is the result of a rational cost-benefit analysis. Most empirical research on the economics of crime aims at testing this hypothesis, which implies, among other things, that economically weaker individuals (*e.g.* unemployed workers) have a higher propensity to commit crime because they face lower opportunity

---

<sup>9</sup>The outcomes of the games at this stage will be comparable to those in the literature where participants typically do not know each other.

<sup>10</sup>I will refer to "crime" in a very broad sense, encompassing various types of offenses, from misdemeanor to felony.

costs. This paper is most closely related to the strand of this literature that focuses on economic factors. For instance, (Kelly, 2000) shows that economic inequality is an important factor of crime, and Gould et al. (2002) and Machin and Meghir (2004) emphasize the role of lower wages (for low-skilled workers) in explaining criminal activities. A plethora of papers have also looked at the role of (un)employment. Using US state data, Raphael and Winter-Ebmer (2001) demonstrate positive effects of unemployment on property crime rates. (Fougère et al., 2009) specifically focus on youth unemployment rate in French regions and show that it increases crime. More recently, Rege et al. (2019) study young displaced male workers during mass layoffs in Norway, and find a strong increase in their propensity to commit a crime in the short run. Apart from this last paper, which relies on individual level data and therefore enables to study individuals' actual behavior, existing literature usually builds data aggregated over large spatial units and thus ignore heterogeneity within their unit of analysis.

By contrast, I argue that according to the type of crime and the expected pay-off, criminals might either operate in their own neighborhood or travel to a remote area, which cannot be captured using aggregate data. It is for instance reasonable to think that thieves are more likely to live in deprived neighborhoods and to steal from wealthier (possibly neighboring) areas, while young delinquents do not have any incentives to bear transportation costs in order to vandalize cars in a distant neighborhood. My paper contributes to the literature by revisiting the question of unemployment's criminogenic effect at the census block level, enabling me to show that unemployment has a differential effect on crime victimization depending on the type of crime and its location. As will be explained below, these results suggests that criminals commit their forfeit in a different location than where they live when they expect a financial loot from their crime.

This paper relies on detailed data from a specific version of the French Victimization Survey (*Cadre de Vie et Sécurité*, INSEE), in which respondents' census block of residence is identified.<sup>11</sup> This survey present several additional valuable features. First, in some cases, it is possible to know where the victimization took place, and hence to control for the characteristics of this location. By contrast, studies relying on police data usually consider the location of the police station where the crime was reported rather than the location of the event itself. Second, these data provide detailed information on individuals, so that relevant individual characteristics pertaining to potential victims' attractiveness can be taken into account, while they are ignored

---

<sup>11</sup>The geographic unit considered, called IRIS, is a neighborhood made of 2,000 individuals on average. Importantly, it is the smallest spatial unit for which representative social and economics indicators can be constructed from census data in France.

in most of the existing literature. Finally, victimization surveys are known for avoiding the under-reporting issue from which reported police data suffer. This data is therefore particularly insightful regarding petty crimes and assaults, which turn out to be driven by different channels than standard financially-motivated crimes. Importantly, I am able to match the blocks where surveyed individuals live to Census data, so that I can characterize the social and economic conditions of these blocks. In this paper, the outcome I study is the individual probability to be victim of a crime, as reported in the survey, and I seek to explain how it is affected by the local unemployment rate.

This study is not exempt from the concern of endogenous sorting that was already raised in section 2.1.2 that would lead to biased estimates of the effect of unemployment on victimization. In order to address this issue, I rely once more on the fixed effect method initiated by Bayer et al. (2008a). In the context of this paper, I use the partition of the French territory provided by the census bureau: municipalities are divided into Large Neighborhoods (*Grands Quartiers*) that encompass several Census Tracts (*IRIS*). I argue that households may select the Large Neighborhood in which they live, but are unlikely to choose the precise Census Tract where they actually reside. I propose some statistical tests to support this assumption. This hypothesis implies that once we control for the characteristics of the Large Neighborhood selected by the household, the remaining spatial variation of unemployment rate across census tracts within the larger area is plausibly exogenous. Using this approach, I confirm that local unemployment rate (i.e. census tract level) is a strong predictor of an individual's probability to be victim of a crime in her neighborhood.

This paper also sheds new light on the mechanisms behind this relationship, by looking at the unemployment rate of both the reference neighborhood, and the first and second rings of adjacent neighborhoods. This enables me to account for heterogeneity across neighborhoods. The results reveal that for burglaries and thefts of objects from cars, unemployment rate in the adjacent neighborhoods have a stronger explanatory power than unemployment in the precise neighborhood where the misdeed occurred. On the contrary, local unemployment rate dominates over more distant unemployment rates in explaining vandalism and assaults in particular. A natural interpretation of these findings is that criminals are mobile for economic crimes but not for violent crimes. In other words, they can afford some transportation cost when they expect a financial reward from their mischief, in line with the Beckerian theory of crime. On the other hand, violent crimes and vandalism escape from this logic and relate more to the social disorganization theory (Shaw and McKay,

1942). This spatial approach helps understanding more precisely criminal behavior according to the different types of crimes, and is therefore an important contribution to the literature.

### **2.2.2 Economic conditions, local public finance and electoral outcomes - *Future project***

As a follow-up of the projects presented above, Clément Malgouyres and I plan to explore how economic conditions can affect social cohesion, with a particular attention paid to understanding whether local public spending (e.g. investment in local amenities and public goods) can play a mediating role. We propose to explore this complex question by investigating more specifically the impact of local economic shocks (e.g. related to firms' exposure to international trade opening) on voting behavior. We will then test whether this effect can be attenuated by the response of local public good provision. As far as electoral outcomes are concerned, we will be interested in participation to local and national elections, but also in the electoral support for the far-right party (Front National). The rise of far-right populist party and their economic origins has been a topic of rising interest among economists (see for instance [Malgouyres, 2017](#); [Colantone and Stanig, 2018](#)). In order to isolate the exogeneous component of the local evolution of employment and productivity, we will rely on an instrumental variable estimation using the predicted sectoral composition and wage inequalities based on their past local values and on their national-wide variation over time as an instrument (Bartik-type). Previous work by C. Malgouyres, gives us confidence in the reliability of this type of instrument. After testing the impact different types of local economic shocks on electoral outcomes, we will carry out an instrumental variable mediation analysis in the spirit of [Dippel et al. \(2017\)](#) in order to unpack the overall effect of economic shocks on voting behavior. In particular, we will determine to what extent the adjustment in local public good provision is responsible for the estimated political responses. The analysis will be all the more interesting that recent reform in the local business tax in France is likely to have affected the way local public resources depend on local economic activity ([Charlot et al., 2016](#)). Interestingly, investigating whether local amenities and public goods can act as a buffer between economic conditions and social cohesion would bridge my two main research axis.

### **2.3 Institutional environment, economic conditions and social cohesion**

I now turn to a slightly different question: the role of the institutional environment in influencing social cohesion. In [Michel and Hémet \(2019\)](#), we look at a specific aspect of this question by exploring the effect

of custodial sanctions (e.g. incarceration) relative to non-custodial ones on crime and employment outcomes of drunk-driving offenders. Driving under the influence of alcohol (DUI) is a major risk factor for road traffic accidents and, as such, represents an important public health issue in most countries. In the European Union, 25% of all road fatalities were due to alcohol in 2015, representing around 6,400 fatalities (European Commission, 2015). The same year, this share was estimated at 28%, representing 10,265 fatalities in the United States (NHTSA, 2017). Understanding how to deter this type of deviant behavior is thus essential to reduce the costs associated with road traffic accidents.

Following the seminal work by [Becker \(1968\)](#), several studies have provided empirical evidence of the deterrence effect of sanctions, generally showing that harsher sanctions indeed decrease criminal behavior (see for instance [Helland and Tabarrok, 2007](#); [Drago et al., 2009](#); [Abrams, 2012](#)).<sup>12</sup> However, these findings may not apply to the case of DUI offenders since this type of crime often involves an addiction problem that makes them fall beyond the rational economic theory of crime. This peculiar aspect of drink-driving crimes calls for a specific analysis, for which the economics literature remains scarce. A notable exception is [Hansen \(2015\)](#) who exploits discontinuities in the severity of the punishment based on blood alcohol content thresholds to measure the impact of incarceration on recidivism. Using data from the State of Washington (US), he finds evidence that more severe sanctions reduce the likelihood that a defendant recidivates.

By contrast, we propose to investigate the role of sanctions of different nature: incarceration versus probation. This enables us to take a stand on the *type* of sanction that would be the most effective, rather than on the *severity* of a given type of sanction. This is particularly relevant in the current context of widespread overcrowding in prisons and of public expenditures reduction, from which the judicial system is not exempt. In this respect, this paper also echoes the small literature on electronic monitoring ([Di Tella and Schargrodsky, 2013](#); [Henneguelle et al., 2016](#)).

To this aim, we take advantage of a large-scale reform of the Danish legislation implemented in 2000, whereby custodial sentences of no more than 60 days were replaced by non-custodial ones: a two-year probation period and a fine, combined with either mandatory participation in a yearlong rehabilitation program or community service.<sup>13</sup> This setting would call for an identification strategy based on a comparison of of-

---

<sup>12</sup>See [Chalfin and McCrary \(2017\)](#) for a recent review.

<sup>13</sup>A few details on the Danish institutional setting are in order here. Prior to the reform, in 1999, 71.9% of the individuals tried for an alleged drunk-driving crime received an unconditional prison sentence. In 95.3% of these cases, the length of the unconditional prison sentence remained below 60 days. Fines and suspensions of the driving license were also frequently used to sanction drink-

fenders whose trial took place before and after the reform, but we show that significant selection occurred in the nature of the offenders tried around the time it was enacted. Indeed, we find evidence that an important share of individuals arrested prior to the reform managed to postpone their trial until after the reform was implemented, so as to benefit from the new law: we observe a sharp drop in the number of cases tried from the moment the law was signed (but before it actually entered into force), as well as a linear decrease in the share of defendants receiving a custodial sentence. Importantly, we also show that wealthier defendants were more likely to have their trial delayed and to avoid incarceration. Incidentally, this finding questions the degree of fairness with which cases can be handled in times of legislative changes, thus adding to the growing literature documenting sources of dysfunction in justice systems (see for instance recent papers by [Philippe and Ouss, 2018](#); [Cohen and Yang, 2019](#)).

This important selection problem precludes us from comparing individuals tried before and after the reform in order to identify the differential effect of incarceration. We circumvent this issue by proposing a novel instrumental variable for the probability of being incarcerated. We use variation in offenders' probability of being incarcerated depending on the date of their crime as a source of identification (an approach loosely related to the one used in [Drago et al., 2009](#)). More precisely, our instrument is the time lapse between the date the crime was committed and the date the reform was enacted. The idea behind this instrument is that the closer to the reform a crime was committed, the more likely the defendant was to be tried under the new law and placed on probation instead of being incarcerated. We document that there is indeed significant variation in the probability of being tried after the reform among offenders arrested up to one year before the reform. In addition, the variation in this variable entirely relies on the date of occurrence of the crime, which we argue to be plausibly exogenous.

Using the time passed between the occurrence of the crime and the date of the reform as an instrument, we are able to measure the differential impact of a custodial sentence relative to a non-custodial one on several crime and labor market outcomes. We do not find any significant effect on recidivism for similar drink-driving crimes, ruling out any deterrence effect of incarceration for such type of crimes. However, we find that, compared to a non-custodial sentence, being incarcerated increases the average number of other drivers, as they were requested in about a third of the cases. Conversely, conditional prison sentences (probation) and community work were either rarely used or not used at all to sanction drunk-drivers. Offenders who received an unconditional prison sentence usually served their time in one of the country's open prisons (low-security prisons where fences, walls, and barriers are minimal, offering softer detention conditions for the inmates).

crimes that offenders commit afterwards. This suggests that the criminogenic effect of custodial sentences impinge on a broad set of offenders, including subsets who exhibit relatively low proclivity for criminal behaviors such as drink-drivers.

Moreover, we find that, compared to non-custodial sentences, custodial ones significantly weaken offenders' labor market attachment as they reduce both their annual number of days worked and earnings. Strikingly, the labor market effects arise more rapidly than the crime effects (around 3 years after trial for the former, versus more than 5 years after for the latter), indicating a potential channel driving subsequent involvement in criminal activities. This last set of results tie our paper to the large and increasing literature on the economic and social consequences of incarceration (see recent studies by [Dobbie et al., 2018](#); [Manudeep et al., 2020](#); [Michel et al., 2019](#)).

### **3 Urban environment and socio-economic outcomes**

My second research axis aims at investigating whether urban amenities and infrastructure affect social cohesion and economic outcomes. As in the first research axis, my research addresses various aspects of this broad question. My joint work on the role of transportation infrastructure on the location and growth of population and economic activity can be seen as a first step into this research axis. Indeed, this set of papers contributes to understand how such public investment influences workers' location as well as the spatial shape of economic activity, which is part of the question at hand. These papers are presented in section 3.1. Another part of my research remains directed toward social cohesion, and in particular crime. The projects presented in section 3.2 aim at understanding how this specific aspect of horizontal social cohesion can be influenced by urban amenities and infrastructure.

#### **3.1 From transportation infrastructure to metropolitan employment**

In a set of papers joint with Miquel-Àngel Garcia-López (Universitat Autònoma de Barcelona and Institut d'Economia de Barcelona) and Elisabet Viladecans-Marsal (Universitat de Barcelona and Institut d'Economia de Barcelona), published in the *Journal of Regional Science* and in *Regional Science and Urban Economics*, we investigate the impact of transportation infrastructure on the location and local growth of population and employment, and on the shape of the city ([Garcia-López et al., 2017a,b](#)). These papers

constitute a first step in understanding how urban infrastructure influence economic outcomes.

The urban economics literature has a long tradition of seeking to explain the determinants of city structure and growth. This literature provides alternative explanations so as to why some cities are more successful and grow more rapidly than others, and how the urban structure changes: while some authors emphasize the importance of human capital and skills (Moretti, 2004; Rosenthal and Strange, 2008, are good examples), others focus on the role of the weather (Glaeser et al., 2001) or the availability of consumer amenities (Carlino and Saiz, 2008) as factors that attract population. Ultimately, however, the literature agrees that a city's density largely accounts for its capacity to be productive and to attract better firms and workers (Combes et al., 2012). For urban economists, agglomeration economies are therefore considered an important source of city growth. What is also clear is that the advantages provided by agglomeration economies increase with a reduction in the transportation costs for goods and people (Glaeser and Gottlieb, 2009). Yet, even though transportation seems to be a key element for employment and population distribution and growth, its impact on urban growth has long been ignored in the empirical literature (see the pioneering work by Duranton and Turner, 2012).

In our two papers, we contribute to this relatively recent branch of the literature by analyzing the effects of a major improvement in transportation infrastructure in the Paris metropolitan area, the Regional Express Rail (RER henceforth), on complementary outcomes: in Garcia-López et al. (2017a), we focus on employment and population growth in the treated municipalities, before turning to the overall city structure (expressed in terms of employment centers' location) in Garcia-López et al. (2017b). It is worth noting that while previous studies typically focus on one specific type of transportation infrastructure, we strive to take into account the various infrastructure that were all improved over the period. Doing so, we provide evidence that one type of train network, the RER, was particularly important for the outcomes that we study. Since both papers use the same setting and rely on similar identification strategies, I now present these common features, before turning to the results specific to each paper.

Both papers focus on the Paris metropolitan area (the *Ile de France* region), which has undergone major road and rail network changes over the past fifty years, and in particular the introduction of a new suburban train in the early 1970s, the *Réseau Express Régional* (RER henceforth). Since its inauguration in 1975, the total length of the RER network has increased by around 550 km, and now operates 257 stations connecting more than 170 municipalities. Metro and tramway networks have also been expanded over the period. Most

notably, the area's main road system (highways) has been extended by 600 km, while the number of ramps and accessibility to many other municipalities have been increased.

In addition to using data on the location and opening date of transportation infrastructure, our work relies on census data between 1968 and 2010 to document the socio-economic characteristics of the 1,300 municipalities in the metropolitan area. In particular, we reveal that metropolitan Paris is undergoing a process of employment decentralization in which its central city (Paris) loses jobs in favor of suburban locations, both in absolute and relative terms. First, the number of jobs in central Paris declined by 7.1% between 1968 and 2010. Second, while the metropolitan area as a whole grew by about one third, the share of jobs located in central Paris dropped from 45.3 to 31.7%. In both papers, we seek to understand the role played by transportation infrastructure in this process.

In this research as well, the main difficulty lies in the identification of an unbiased causal impact of transport infrastructure, which is threatened by the endogenous location of new infrastructure. Indeed, planners may decide to improve the connection of deprived areas in order to boost their economic activity or attract population, or they may on the contrary connect areas that are expected to grow, anticipating on future needs for transportation. Over the past decade, the literature has come up with various strategies to address this issue. A popular method consists in using exogenous deviations from initial transportation plans as a source of identification (see for instance [Baum-Snow, 2007](#); [Michaels, 2008](#); [Mayer and Trévien, 2015](#)). An alternative common strategy is to rely on historical instruments. For instance, [Duranton and Turner \(2012\)](#) instrument road infrastructure using the U.S. railway network at the end of the 19th century and the routes taken by major expeditions of the United States between 1518 and 1850; [Hsu and Zhang \(2014\)](#) rely on the historical railway network plan of 1890 and the planned national express way extension as exogenous sources of variation of highway location in Japan; and [Garcia-López et al. \(2015\)](#) use the Roman roads and the 1760 Postal routes as instruments for Spanish highways. Our empirical strategy follows this approach, as we rely on two historical instruments, the 1870 railways and the Roman roads, as a source of exogenous variation of current infrastructure' location.

The general argument to support the credibility of historical instruments is the following: at least some parts of the current infrastructure are built where the ancient one was located, for two main reasons. First, past infrastructure was built in a specific location to take advantage of favorable physical geography (*e.g.* low

elevation, no slope...), and this usually remains true several centuries later. Second, the old infrastructure can be at least partially re-used in the construction of the new one, thus reducing its costs. In the case of the Paris Metropolitan area, a visual inspection of the 1870 railway network immediately reveals the similarity with the current network of commuter trains in general and the RER in particular.

Moreover, the exclusion restriction imposes that the historical infrastructure (in our case, the route of the 1870 rail network for instance) must not directly influence the outcome under study (for instance, current growth in local employment). Although this hypothesis is not directly testable, it is generally admitted in the literature, particularly because of the important time lag between historical and current infrastructure. Yet, we follow previous studies and include a number of geographical and historical characteristics in our analysis to strengthen the credibility of this hypothesis (otherwise stated, we only guarantee "conditional" exogeneity of our instrument). We thus take into account the characteristics of the terrain such as its altitude, slope or ruggedness, insofar as they can explain both past and present infrastructure. In addition, we take into account the location of former economic centers such as Roman camps, major cities between the 12th and 16th centuries, or former monasteries as this may explain both the presence of transport infrastructure in the past, and current economic outcomes.

In the paper entitled "**How does transportation shape intrametropolitan growth? An answer from the Regional Express Rail**", published in the *Journal of Regional Science* (Garcia-López et al., 2017a) we show that the proximity of a RER station increases employment, population density and growth rate of these two variables. In a first step, we seek to understand to what extent the RER determines employment location. We therefore estimate the effect of the distance to the nearest RER station on employment density for all 1,300 municipalities in the metropolitan area. Our results indicate that if an RER station were built one kilometer closer to a given municipality, then employment density would increase by 6.5% in that municipality. The estimated effect is slightly lower (in the order of 5.5%) when distance to other transport infrastructure (other commuter train or access to a high-speed road) is taken into account. Importantly, although other transportation infrastructure have significant effects, including them does not significantly reduce the estimated RER effect.

In a second step, we look at the effect of the RER on municipal employment growth, in order to better understand the dynamic effects of this type of infrastructure on the local economy. To do so, we estimate

the effect of distance to the nearest RER station on the variation in employment density between 1968 and 2010. Our results reveal that moving away from an RER station tends to reduce employment growth. More precisely, the coefficients tell us that, on average, each additional kilometer separating a city from the nearest RER station reduces its employment growth by 2%, whether we control for the proximity to another type of transport infrastructure or not.

Finally, we analyze the heterogeneity of these results along several dimensions. In terms of spatial heterogeneity, we observe that the effect of the RER on employment growth is larger for municipalities which are the closest to a RER station and the closest to an employment (sub)center. Turning to heterogeneity in terms of employment characteristics, we find that the growth effect is stronger for jobs in the service sector and for population with high school and university degrees in particular. Eventually, looking at the heterogeneity over time reveals that there is no impact of the RER expansion on employment growth during the first part of the covered period suggesting some lagged effect of such massive investment.<sup>14</sup>

The companion paper entitled "**Next Train to the Polycentric City: The effect of railroads on subcenter formation**", published in *Regional Science and Urban Economics* (Garcia-López et al., 2017b) focuses on the role of transportation infrastructure in the structure of the city. We start by establishing that the decentralization process documented above follows a polycentric spatial pattern, whereby existing employment subcenters are reinforced, and new subcenters emerge.<sup>15</sup>

After analyzing the job decentralization process in the Paris metropolitan area, we establish that rail transit causes subcenter formation, which is the main contribution of this paper. To answer to this key question, we proceed in two steps. We first investigate whether the existence of a rail station in a suburban municipality increases the probability that this municipality becomes (part of) a subcenter. Then, we examine whether proximity to rail stations also increases the likelihood of becoming (part of) a subcenter, even when the station is not built on the municipal ground.

Our results reveal that on average, the probability of a suburban municipality of being (part of a) subcenter increases by 5% if the municipality has a rail station within its boundaries, or by 3% if, although not having a railstation, municipality proximity to a rail station increases by 10%. Results for the RER confirm its effect is stronger: a 10% and a 5% increase in the probability of becoming (part of a) subcenter if the

---

<sup>14</sup>By contrast, the impact on population growth, which are not reported here, was sizable much earlier but declined over time.

<sup>15</sup>This pattern contrasts with dispersed decentralization in which suburban land is occupied by low-density settlements.

municipality has a RER station or if municipality proximity to a RER station increases by 10%, respectively.

### 3.2 From urban amenities and infrastructure to social cohesion

I eventually turn to presenting two projects that seek to understand how urban amenities and infrastructure affect social cohesion, with a specific focus on crime.

**Urban renewal and crime - *Recent ongoing project*** In a new ongoing project joint with Nina Guyon (National University of Singapore) and Arnaud Philippe (University of Bristol), our purpose is to explore how the urban environment influence crime. Several studies have been devoted to the effect of the institutional and urban environment on crime and the feeling of insecurity, in particular following Wilson and Keling's controversial article on the broken window theory.<sup>16</sup> This theory states that a deteriorated urban environment is an external sign of low social cohesion and of weak policing activities that would lead to a higher level of crime. More recent academic literature has documented the effects of the social and urban environment on crime or insecurity. Relying on the "Moving To Opportunity" experiment, in which randomly selected deprived households were given a chance to move to a better neighborhood, [Ludwig et al. \(2001\)](#) suggest that this program led to lower child delinquency. Relatedly, [Freedman and Owens \(2011\)](#) focus on the effect of urban renovations of social housing on crime, and find a decrease in acts of violence but no effect on damage to property.

In this project, we measure the crime effects of a massive urban renewal program in France (PNRU henceforth) initiated in 2004, whereby the French government has invested 47 billion euros to subsidize the demolition and reconstruction of housing (mainly social housing), as well as improvements to the quality of life (green spaces, outdoor facilities, etc.) in 571 Sensitive Urban Zones. To this purpose, we combine the detailed financial data for the PNRU over the 2004-2015 period<sup>17</sup> to the administrative data of reported crimes in each police district since 1995<sup>18,19</sup>. Interestingly, because the PNRU involves both population displacement related to demolition and reconstruction of social housing units and improvement in the quality

---

<sup>16</sup>The article entitled "Broken windows" was published in *The Atlantic* in March 1982.

<sup>17</sup>AGORA data, French Agency for Urban Renewal (ANRU) and the Commissariat Général à l'Égalité des Territoires (CGET)

<sup>18</sup>Etat 4001, French Ministry of Interior

<sup>19</sup> Although police districts are much larger than the neighborhoods targeted by the urban regeneration policy, we make this an asset by claiming that this limits the risk that our results are driven by crime displacement. Indeed, an important part of treated housing blocks' population is displaced, at least in the short run, so that looking at outcomes at the neighborhood level would be misleading.

of local public spaces and amenities, this program might enable us to disentangle the "social" effect from the "physical environment" effect of such policies. In the medium run, we also plan to look at the effect of the PNRU on election turnout and share of votes obtained by the main extreme-right party (former National Front), considered as alternative indicators of social cohesion.

In terms of identification of the effect, this policy offers a very neat setting to study the causal impact of various types of urban infrastructure and amenities on social cohesion because the targeted neighborhoods were a subsample of a larger set of deprived neighborhoods (called Sensitive Urban Zones, ZUS in French), thus providing us with a treated and a control group that are very comparable in some aspects. We are currently working on two alternative identification strategies. First, we develop a difference-in-difference strategy where we use non-treated ZUS as a control group, following [Guyon \(2017\)](#) and subsequent work. More precisely, police districts that include sensitive urban zones which were part of the urban renewal program are considered as treated, while police districts that include sensitive urban zones which were *not* part of the program are used as control areas. Alternatively, we are relying on an event-study analysis, where different "events" can be considered: the date of signature of the project in each neighborhood and the starting date of the first renovation work are possible candidates. Incidentally, this project relates to a series of papers studying the effect of the French enterprise zones program, whereby tax exemptions were offered to firms settling in some of these Sensitive Urban Zones (see [Gobillon et al., 2012](#); [Briant et al., 2015](#); [Mayer et al., 2015](#), for three excellent studies).

In the medium run, we also plan to look at the effect of this program on election turnout and share of votes obtained by the main extreme-right party (former National Front), considered as alternative indicators of social cohesion.

**Transportation infrastructure and crime - *Future project*** Eventually, a natural follow-up of the projects presented above will be to investigate the impact of transportation infrastructure on crime. I first plan to focus on the Paris Metropolitan area, for which data about the location and construction date of suburban train stations and high speed roads are readily available (since the early 1970s), before enlarging the scope of the analysis depending on the availability of other data. In order to address the problem of endogenous location of transportation infrastructure, I will rely on historical instruments following [Garcia-López et al. \(2017a\)](#) and [Garcia-López et al. \(2017b\)](#) to study social cohesion outcomes, that are even less likely to be directly

influenced by historical transportation infrastructure than employment and population growth. Regarding the most recent railway network in the region (RER), an alternative source of identification could be to use the fact that the current network exogenously deviated from the initially planned network, as in [Mayer and Trevien \(2017\)](#). Crime being particularly prone to displacement effects, it will be especially relevant to account for spillover effect in this context. Thus, one must not only assess the effect of the presence of a transportation infrastructure (e.g. train station) in the area where it is located (the reference area), but also in areas to which it connects the reference area. To this aim, I plan to rely on exogenous variation in the commuting time between two locations related to strikes, important breakdowns or major works requiring the closure of a station, in the spirit of [Ridgeway and MacDonald \(2016\)](#). This will be made possible for more recent periods as daily crime data can be available from the Ministry of Interior, as well as precise dates of opening and closing of stations from the Transilien SNCF company.

## 4 Research-related activities

### 4.1 Masters' students supervision and related activities

Since I started working as an assistant professor at Paris School of Economics, I have **supervised the Masters' dissertations** of two students enrolled in the program "Analysis and Policy in Economics". In 2016-2017, I supervised **Marie Gaudier**, on a project entitled "Is Crime an Environmental Phenomenon? The Role of Local Equipments on Crime Occurrence in French Municipalities", in which she explored whether urban public infrastructure and equipment have an impact of local crime rates. After graduating from the master's program, Marie Gaudier was hired as a consulting analyst at Veltys, a strategy consulting firm specializing in data analysis and modeling. In 2017-2018, I supervised **Madeleine Péron** on a project entitled "Does Unemployment Cause Crime?", where she explored the relationship between unemployment and crime using recently available French data. At the end of her master's program, Madeleine Péron was hired as a research assistant at the *Observatoire du Bien-être*, CEPREMAP.

I was also a **referee for several masters' dissertations** at PSE: **Georgia Thébaud**, supervised by Julien Grenet (PSE) on "Geographical constraint and selective programs in French higher education." in 2018; **Julien Deur**, supervised by Laurent Gobillon (PSE) on "How do real estate prices react to a gain in accessi-

bility ? A case study in the Paris region." in 2018; and **Elika Attari**, supervised by Gregory Verdugo (Evry), on "Immigrants Occupational Specialization in Europe Evidence from 2005 to 2014" in 2017.

Since 2018, together with François Fontaine (PSE), I have been **in charge of a research seminar** during which PSE masters' students discuss papers related to their research topic, before presenting the first elements of their master's dissertations: problems, methods considered... The objective of this seminar is to help students clarify their research question before they actually start their master's thesis and to guide them towards relevant methods, to ensure that their research starts under the right conditions.

## 4.2 Experience related to PhD students supervision

Since I defended my PhD in November 2013, I have been a **member of several PhD defense juries**: **Simón Planells Struse** (University of Barcelona) defended a thesis entitled "Essays on the Economics of Crime: Determinants of Crime in an Urban Context", supervised by Daniel Montolio Estivill, in 2015; **Andrea Aguirre Sanchez** (Saint-Etienne University), defended a thesis entitled "Urban Crime in Ecuador: Three essays on the role of economic inequalities, population density and emotions" supervised by Nelly Exbrayat et Stéphane Riou, in 2018; **Gudrun Johnsen** (École Normale Supérieure de Paris) will defend a thesis entitled "Risk taking at the corporate level: In search of sustainable growth" supervised by Gilles Saint-Paul in December 2019; and **Mariona Segú** (Paris Sud University) will defend a thesis on housing economics supervised by Miren Lafourcade and Gabrielle Fack in 2020.

I have also **participated in several PhD thesis committees** (the objective of these committees is to give the students some feedback on their ongoing work about a year before their defense, so as to guide them towards the next steps of their dissertation): **Andrea Aguirre Sanchez** (Saint-Etienne University) in 2017, **Benjamin Carantino** (PSE) in 2018 (supervised by Laurent Gobillon and Miren Lafourcade) and **Gudrun Johnsen** (ENS-PSL) in 2019.

Since 2016, I **contribute actively to the job-market training** of final-year PhD students at PSE. In particular, I regularly conduct mock interviews for students who are on the international job market. Moreover, I present the hiring process on the French job market for assistant professors (Maîtres de Conférences) during an annual information session for PhD students.

### 4.3 Contribution to scientific projects and events

In 2018, the French Research Agency (ANR) granted me funding for a project that I proposed as **Principal Investigator** (ANR-18-CE22-0013-01; 165,000 €): "Social cohesion and the city: Urban infrastructure and amenities for a sustainable society". This funding should allow me to finance the ongoing research and future projects described above in this dissertation. In particular, this funding enables me to coordinate a team of young researchers, including a PhD student and a post-doctoral student who was hired thanks to this grant.

Before that, **I participated to several research projects as researcher**. I took part in the project entitled "Housing and salary: a joint analysis of two key dimensions of immigrant and refugee assimilation", funded by the ANR (ANR-16-FASI-0001-01; 199,800 €), and coordinated by Gregory Verdugo (University of Evry) between 2016 and 2019. I was also a member of two other projects coordinated by E. Viladecans-Marsal (UB, IEB) between 2014 and 2016: "Global challenges, local responses: the role of cities as engine of economic prosperity" funded by the Spanish Ministry of Economy Ministerio de Economía y Competitividad (ECO-2013-41310-R; 20,300 €); and "Infrastructures' Improvement and Urban Structure: the Case of Paris", funded by the *Société du Grand Paris* (CRC-2013-06; 34,000 €).

On another note, I have been regularly involved in the **organization of seminars and workshops**. During my post-doctoral experience at the Institute of Economics of Barcelona, I **co-organized the international seminar**. Soon after I started working at PSE, I took over the organization of the international and internal seminars of the Labor and Public Economics research group, which I am still co-organizing. When I was a post-doc in Barcelona, I also contributed to **organize the Third IEB Workshop in Urban Economics** in 2014. Since then, I have been **member of scientific committees for several workshops**: 8<sup>th</sup> European Meeting of the Urban Economics Association (UEA) and 67<sup>th</sup> meeting of the French Economic Association in 2018; 9<sup>th</sup> European Meeting of the UEA in 2019.

### 4.4 Other administrative and collective responsibilities

Since 2017, I was a **member of several hiring committees for assistant professors (maîtres de conférences) in French universities**: University of Grenoble in 2017, École Normale Supérieure de Lyon, Aix-Marseille University, Paris 1 Panthéon-Sorbonne University and Paris-Est-Créteil University in 2018;

and Aix-Marseille University in 2018.

Since September 2017, I have been **Head of Studies of the Economics department at ENS**. My main role consists in counseling ENS students with respect to their academic track choices. In particular I advise them when they write research proposals in order to apply for PhD grants.

On another note, since January 2019, I have been an **elected member of PSE's scientific council**, where issues such as PSE hiring policy or scientific orientation are discussed and decided upon.

## 5 Bibliography

- Abrams, D. S. (2012). Estimating the deterrent effect of incarceration using sentencing enhancements. *American Economic Journal: Applied Economics*, 4(4):32–56.
- Albouy, D. (2016). What are cities worth? land rents, local productivity, and the total value of amenities. *The Review of Economics and Statistics*, 98(3):477–487.
- Alesina, A., Baqir, R., and Easterly, W. (1999). Public goods and ethnic divisions. *The Quarterly Journal of Economics*, 114(4):1243–1284.
- Alesina, A. and La Ferrara, E. (2000). Participation in heterogeneous communities. *The Quarterly Journal of Economics*, 115(3):847–904.
- Alesina, A. and La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics*, 85(2):207–234.
- Alesina, A. and La Ferrara, E. (2005). Ethnic diversity and economic performance. *Journal of Economic Literature*, 43(3):762–800.
- Algan, Y. and Cahuc, P. (2010). Inherited trust and growth. *The American Economic Review*, 100(5):2060–2092.
- Algan, Y. and Cahuc, P. (2014). Trust, Growth, and Well-Being: New Evidence and Policy Implications. In *Handbook of Economic Growth*, volume 2 of *Handbook of Economic Growth*, chapter 2, pages 49–120. Elsevier.
- Algan, Y., Hémet, C., and Laitin, D. D. (2016). The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation. *Journal of Political Economy*, 124(3):696–733.
- Altonji, J. G. and Mansfield, R. K. (2018). Estimating Group Effects Using Averages of Observables to Control for Sorting on Unobservables: School and Neighborhood Effects. *American Economic Review*, 108(10):2902–2946.
- Battu, H., Seaman, P., and Zenou, Y. (2011). Job contact networks and the ethnic minorities. *Labour Economics*, 18(1):48–56.
- Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805.
- Bayer, P., Ross, S. L., and Topa, G. (2008a). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy*, 116(6):1150–1196.

- Bayer, P., Ross, S. L., and Topa, G. (2008b). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy*, 116(6):1150–1196.
- Beauvais, C., Sc, M., and Jenson, J. (2002). Social cohesion: Updating the state of the research. Cprn discussion paper no. fl22 (canadian policy research networks, ottawa).
- Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76:169–217.
- Berge, L. I. O., Bjorvatn, K., Galle, S., Miguel, E., Posner, D. N., Tungodden, B., and Zhang, K. (2019). Ethnically Biased? Experimental Evidence from Kenya. *Journal of the European Economic Association*. jvz003.
- Bertrand, M., Luttmer, E. F. P., and Mullainathan, S. (2000). Network effects and welfare cultures. *The Quarterly Journal of Economics*, 115(3):1019–1055.
- Bifulco, R., Fletcher, J. M., and Ross, S. L. (2011). The effect of classmate characteristics on post-secondary outcomes: Evidence from the add health. *American Economic Journal: Economic Policy*, 3(1):25–53.
- Boucher, V., Bramoullé, Y., Djebbari, H., and Fortin, B. (2014). Do peers affect student achievement? evidence from canada using group size variation. *Journal of Applied Econometrics*, 29(1):91–109.
- Bramoullé, Y., Currarini, S., Jackson, M. O., Pin, P., and Rogers, B. W. (2012). Homophily and long-run integration in social networks. *Journal of Economic Theory*, 147(5):1754–1786.
- Briant, A., Lafourcade, M., and Schmutz, B. (2015). Can tax breaks beat geography? lessons from the french enterprise zone experience. *American Economic Journal: Economic Policy*, 7(2):88–124.
- Calvó-Armengoi, A. and Zenou, Y. (2004). Social networks and crime decisions: The role of social structure in facilitating delinquent behavior. *International Economic Review*, 45(3):939–958.
- Calvó-Armengol, A. and Jackson, M. O. (2004). The effects of social networks on employment and inequality. *American Economic Review*, 94(3):426–454.
- Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). Peer Effects and Social Networks in Education. *The Review of Economic Studies*, 76(4):1239–1267.
- Calvó-Armengol, A., Verdier, T., and Zenou, Y. (2007). Strong and weak ties in employment and crime. *Journal of Public Economics*, 91(1-2):203–233.
- Card, D. (2001). Immigrant inflows, native outflows, and the local labor market impacts of higher immigration. *Journal of Labor Economics*, 19(1):22–64.
- Carlino, G. A. and Saiz, A. (2008). City beautiful. Working Paper 08–22, Federal Reserve Bank of Philadelphia.
- Carrell, S., Hoekstra, M., and West, J. (2020). The Impact of College Diversity on Behavior toward Minorities. *American Economic Journal: Economic Policy*.
- Chalfin, A. and McCrary, J. (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature*, 55(1):5–48.

- Chan, J., To, H.-P., and Chan, E. (2006). Reconsidering Social Cohesion: Developing a Definition and Analytical Framework for Empirical Research. *Social Indicators Research*, 75(2):273–302.
- Charlot, S., Paty, S., and Riou, S. (2016). From competition to cooperation: on the spatial effects of a local business tax policy reform in France. Post-Print halshs-01421041, HAL.
- Cohen, A. and Yang, C. S. (2019). Judicial politics and sentencing decisions. *American Economic Journal: Economic Policy*, 11(1):160–191.
- Colantone, I. and Stanig, P. (2018). The trade origins of economic nationalism: Import competition and voting behavior in western europe. *American Journal of Political Science*, 62(4):936–953.
- Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(Supplement: Organizations and Institutions:Sociological and Economic Approaches to the Analysis of Social Structure):S95–S120.
- Combes, P.-P., Duranton, G., and Gobillon, L. (2018). The Costs of Agglomeration: House and Land Prices in French Cities. *The Review of Economic Studies*, 86(4):1556–1589.
- Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6):2543–2594.
- Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *The American Economic Review*, 100(1):35–69.
- Corcoran, M., Datcher, L., and Duncan, G. (1980). Information and influence networks in labor markets. *Five Thousand American Families: Patterns of Economic Progress*, 7:1–37.
- Currarini, S., Jackson, M. O., and Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045.
- Damm, A. P. and Dustmann, C. (2014). Does growing up in a high crime neighborhood affect youth criminal behavior? *The American Economic Review*, 104(6):1806–1832.
- De Giorgi, G., Frederiksen, A., and Pistaferri, L. (2020). Consumption network effects. *Review of Economic Studies*.
- Di Tella, R. and Schargrodsy, E. (2013). Criminal recidivism after prison and electronic monitoring. *Journal of Political Economy*, 121(1):28–73.
- Dippel, C., Gold, R., Heblich, S., and Pinto, R. (2017). Instrumental variables and causal mechanisms: Unpacking the effect of trade on workers and voters. Working Paper 23209, National Bureau of Economic Research.
- Dobbie, W., Gronqvist, H., Niknami, S., Palme, M., and Priks, M. (2018). The intergenerational effects of parental incarceration. Working Paper 24186, National Bureau of Economic Research.
- Drago, F., Galbiati, R., and Vertova, P. (2009). The deterrent effects of prison: Evidence from a natural experiment. *Journal of Political Economy*, 117(2):257–280.
- Duranton, G. and Turner, M. A. (2012). Urban growth and transportation. *The Review of Economic Studies*, 79(4):1407–1440.

- Dustmann, C., Glitz, A., Schönberg, U., and Brücker, H. (2016). Referral-based Job Search Networks. *The Review of Economic Studies*, 83(2):514–546.
- Easterly, W., Ritzen, J., and Woolcock, M. (2006). Social cohesion, institutions and growth. *Economics & Politics*, 18(2):103–120.
- Edward L. Glaeser, J. K. and Saiz, A. (2001). Consumer city. *Journal of Economic Geography*, 1(1):27–50.
- Ehrlich, I. (1973). Participation in illegitimate activities: A theoretical and empirical investigation. *Journal of Political Economy*, 81(3):521–65.
- Fearon, J. D. and Laitin, D. D. (1996). Explaining interethnic cooperation. *American Political Science Review*, 90(4):715–735.
- Fearon, J. D. and Laitin, D. D. (2000). Violence and the social construction of ethnic identity. *International Organization*, 54(04):845–877.
- Feler, L. and Senses, M. Z. (2017). Trade shocks and the provision of local public goods. *American Economic Journal: Economic Policy*, 9(4):101–43.
- Fershtman, C. and Gneezy, U. (2001). Discrimination in a Segmented Society: An Experimental Approach\*. *The Quarterly Journal of Economics*, 116(1):351–377.
- Finseraas, H. and Kotsadam, A. (2017). Does personal contact with ethnic minorities affect anti-immigrant sentiments? evidence from a field experiment. *European Journal of Political Research*, 56(3):703–722.
- Fletcher, J. M. (2010). Social interactions and smoking: evidence using multiple student cohorts, instrumental variables, and school fixed effects. *Health Economics*, 19(4):466–484.
- Fougère, D., Kramarz, F., and Pouget, J. (2009). Youth unemployment and crime in france. *Journal of the European Economic Association*, 7(5):909–938.
- Francois, P. and Zabojnik, J. (2005). Trust, social capital, and economic development. *Journal of the European Economic Association*, 3(1):51–94.
- Freedman, M. and Owens, E. G. (2011). Low-income housing development and crime. *Journal of Urban Economics*, 70(2):115 – 131.
- Fricke, H., Grogger, J., and Steinmayr, A. (2018). Exposure to academic fields and college major choice. *Economics of Education Review*, 64(C):199–213.
- Gaigné, C., Koster, H. R., Moizeau, F., and Thisse, J.-F. (2017). Amenities and the Social Structure of Cities. CEPR Discussion Papers 11958, C.E.P.R. Discussion Papers.
- Garcia-López, M.-A., Holl, A., and Viladecans-Marsal, E. (2015). Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *Journal of Urban Economics*, 85(0):52 – 67.
- Garcia-López, M.-A., Hémet, C., and Viladecans-Marsal, E. (2017a). How does transportation shape intrametropolitan growth? An answer from the Regional Express Rail. *Journal of Regional Science*, 57(5):758–780.
- Garcia-López, M.-A., Hémet, C., and Viladecans-Marsal, E. (2017b). Next train to the polycentric city: The effect of railroads on subcenter formation. *Regional Science and Urban Economics*, 67(C):50–63.

- Giorgi, G. D., Pellizzari, M., and Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, 2(2):241–275.
- Glaeser, E. L. and Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature*, 47(4):983–1028.
- Glaeser, E. L., Kolko, J., and Saiz, A. (2001). Consumer city. *Journal of Economic Geography*, 1(1):27–50.
- Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996). Crime and social interactions. *The Quarterly Journal of Economics*, 111(2):507–48.
- Gobillon, L., Magnac, T., and Selod, H. (2012). Do unemployed workers benefit from enterprise zones? the french experience. *Journal of Public Economics*, 96(9):881 – 892.
- Gould, E. D., Weinberg, B. A., and Mustard, D. B. (2002). Crime rates and local labor market opportunities in the united states: 1979-1997. *The Review of Economics and Statistics*, 84(1):45–61.
- Goux, D., Gurgand, M., and Maurin, E. (2017). Adjusting your dreams? high school plans and dropout behaviour. *The Economic Journal*, 127(602):1025–1046.
- Goux, D. and Maurin, E. (2007). Close neighbours matter: Neighbourhood effects on early performance at school. *The Economic Journal*, 117(523):1193–1215.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380.
- Granovetter, M. S. (1995). *Getting a Job: a Study of Contacts and Careers*. Cambridge, MA. Harvard University Press, 2nd edition.
- Guyon, N. (2017). Étude des effets de la rénovation urbaine sur l'évolution du bâti et du peuplement dans les quartiers ciblés entre 2004 et 2013. working paper, LIEPP Policy Report.
- Hansen, B. (2015). Punishment and deterrence: Evidence from drunk driving. *The American Economic Review*, 105(4):1581–1617.
- Hanson, J. and Hillier, B. (1987). The architecture of community: Some new proposals on the social consequences of architectural and planning decisions. *Architecture and Behaviour*, 3(3):251–273.
- Helland, E. and Tabarrok, A. (2007). Does three strikes deter?: A nonparametric estimation. *Journal of Human Resources*, 42(2).
- Hellerstein, J. K., Kutzbach, M. J., and Neumark, D. (2019). Labor market networks and recovery from mass layoffs: Evidence from the great recession period. *Journal of Urban Economics*, page 103192.
- Hellerstein, J. K., McInerney, M., and Neumark, D. (2011). Neighbors and coworkers: The importance of residential labor market networks. *Journal of Labor Economics*, 29(4):659–695.
- Henderson, J. V. (1974). The Sizes and Types of Cities. *American Economic Review*, 64(4):640–656.
- Henneguelle, A., Monnery, B., and Kensey, A. (2016). Better at home than in prison? the effects of electronic monitoring on recidivism in france. *Journal of Law and Economics*, 59(3):629–667.
- Hémet, C. (2017). Ethnic Networks and the Informal Labor Market. *Mimeo*.
- Hémet, C. (2018). Unemployment and Crime Victimization: a Local Approach. *Mimeo*.

- Hémet, C. and Malgouyres, C. (2018). Diversity and Employment Prospects: Neighbors Matter! *Journal of Human Resources*, 53(3):825–858.
- Hémet, C. and Malgouyres, C. (2019). The Strength of Local Referral Networks: Evidence from France. *Mimeo*.
- Holzer, H. J. (1988). Search method use by unemployed youth. *Journal of Labor Economics*, 6(1):1–20.
- Hsu, W.-T. and Zhang, H. (2014). The fundamental law of highway congestion revisited: Evidence from national expressways in Japan. *Journal of Urban Economics*, 81(C):65–76.
- Ioannides, Y. M. and Datcher Loury, L. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, 42(4):1056–1093.
- Jenson, J. (2010). *Defining and Measuring Social Cohesion*.
- Jonsson, J. O. and Mood, C. (2008). Choice by contrast in Swedish schools: How peers' achievement affects educational choice. *Social Forces*, 87(2):741–765.
- Kelly, M. (2000). Inequality and crime. *The Review of Economics and Statistics*, 82(4):530–539.
- Knack, S. and Keefer, P. (1997). Does social capital have an economic payoff? a cross-country investigation. *The Quarterly Journal of Economics*, 112(4):1251–1288.
- Koster, H. R. A., van Ommeren, J. N., and Rietveld, P. (2016). Historic amenities, income and sorting of households. *Journal of Economic Geography*, 16(1):203–236.
- Lazear, E. P. (1999). Globalisation and the market for team-mates. *Economic Journal*, 109(454):C15–40.
- Legeby, A., Berghauer Pont, M., and Marcus, L. (2015). Street interaction and social inclusion. In *Suburban Urbanities: Suburbs and the Life of the High Street*, chapter 9. UCL Press.
- Light, I. (2005). *The Handbook of Economic Sociology*, chapter 28. The Ethnic Economy. Princeton University Press, 2nd edition.
- Light, I. and Rosenstein, C. (1995). *Race, Ethnicity, and Entrepreneurship in Urban America*. Walter deGruyter Inc.
- Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4):825–860.
- Ludwig, J., Duncan, G. J., and Hirschfield, P. (2001). Urban poverty and juvenile crime: Evidence from a randomized housing-mobility experiment. *The Quarterly Journal of Economics*, 116(2):655–679.
- Luttmer, E. F. P. (2001). Group loyalty and the taste for redistribution. *Journal of Political Economy*, 109(3):500–528.
- Machin, S. and Meghir, C. (2004). Crime and economic incentives. *Journal of Human Resources*, 39(4).
- Malgouyres, C. (2017). Trade Shocks and Far-Right Voting: Evidence from French Presidential Elections. RSCAS Working Papers 2017/21, European University Institute.
- Manudeep, B., Dahl, G. B., Loken, K. V., and Mogstad, M. (2020). Incarceration, recidivism and employment. *Journal of Political Economy*.

- Mayer, T., Mayneris, F., and Py, L. (2015). The impact of Urban Enterprise Zones on establishment location decisions and labor market outcomes: evidence from France. *Journal of Economic Geography*, 17(4):709–752.
- Mayer, T. and Trévien, C. (2015). The impacts of urban public transportation: Evidence from the Paris metropolitan area. Working Paper G2015/03, INSEE.
- Mayer, T. and Trevien, C. (2017). The impact of urban public transportation evidence from the paris region. *Journal of Urban Economics*, 102(C):1–21.
- McPherson, M., Smith-Lovin, L., and Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444.
- Mendolia, S., Paloyo, A. R., and Walker, I. (2018). Heterogeneous effects of high school peers on educational outcomes. *Oxford Economic Papers*, 70(3):613–634.
- Merton, R. K. (1938). Social structure and anomie. *American Sociological Review*, 3(5):672–682.
- Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4):683–701.
- Michel, B. and Hémet, C. (2019). Custodial vs. non-custodial punishments: Measuring the impact of an anticipated reform. *Mimeo*.
- Michel, B., Rosholm, M., and Simonsen, M. (2019). Measuring the impact of incarceration on life trajectories. *Mimeo*.
- Miguel, E. (2004). Tribe or nation? nation building and public goods in kenya versus tanzania. *World Politics*, 56:328–362.
- Miguel, E. and Gugerty, M. K. (2005). Ethnic diversity, social sanctions, and public goods in kenya. *Journal of Public Economics*, 89(11-12):2325–2368.
- Montgomery, J. D. (1994). Weak ties, employment, and inequality: An equilibrium analysis. *American Journal of Sociology*, 99(5):1212–1236.
- Moretti, E. (2004). Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review*, 94(3):656–690.
- Patacchini, E., Rainone, E., and Zenou, Y. (2017). Heterogeneous peer effects in education. *Journal of Economic Behavior and Organization*, 134:190 – 227.
- Philippe, A. and Ouss, A. (2018). "no hatred or malice, fear or affection": Media and sentencing. *Journal of Political Economy*, 126(5):2134–2178.
- Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. Simon & Schuster, New York.
- Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century the 2006 johan skytte prize lecture. *Scandinavian Political Studies*, 30(2).
- Raphael, S. and Winter-Ebmer, R. (2001). Identifying the effect of unemployment on crime. *The Journal of Law and Economics*, 44(1):259–283.

- Rappaport, J. (2008). Consumption amenities and city population density. *Regional Science and Urban Economics*, 38(6):533 – 552.
- Rappaport, J. (2009). The increasing importance of quality of life? *Journal of Economic Geography*, 9(6):779–804.
- Rege, M., Skardhamar, T., Telle, K., and Votruba, M. (2019). Job displacement and crime: Evidence from norwegian register data. *Labour Economics*, 61:101761.
- Ridgeway, G. and MacDonald, J. (2016). Effect of rail transit on crime: A study of los angeles from 1988 to 2014. *Journal of Quantitative Criminology*, 33.
- Rosenthal, S. and Ross, S. (2015). Change and persistence in the economic status of neighborhoods and cities. *Handbook of Regional and Urban Economics*, 5:1047–1120.
- Rosenthal, S. and Strange, W. (2004). Evidence on the nature and sources of agglomeration economies. *Handbook of Regional and Urban Economics*, 4:2119–2171.
- Rosenthal, S. and Strange, W. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2):373–389.
- Schmutte, I. M. (2015). Job referral networks and the determination of earnings in local labor markets. *Journal of Labor Economics*, 33(1):1–32.
- Shaw, C. and McKay, H. (1942). *Juvenile Delinquency and Urban Areas*. University of Chicago Press.
- Solignac, M. and Tô, M. (2018). Do workers make good neighbours? the impact of local employment on young male and female entrants to the labour market. *Annals of Economics and Statistics*, (130):167–198.
- Speckesser, S. and Hedges, S. (2017). Peer effects and social influence in post-16 educational choice. National Institute of Economic and Social Research (NIESR) Discussion Papers 483, National Institute of Economic and Social Research.
- Tabellini, G. (2008). The Scope of Cooperation: Values and Incentives. *The Quarterly Journal of Economics*, 123(3):905–950.
- Topa, G. and Zenou, Y. (2015). Neighborhood and network effects. volume 5, chapter Chapter 9, pages 561–624. Elsevier.
- Uslaner, E. M. (2012). *Segregation and Mistrust: Diversity, Isolation, and Social Cohesion*. Cambridge University Press.
- Wahba, J. and Zenou, Y. (2005). Density, social networks and job search methods: Theory and application to egypt. *Journal of Development Economics*, 78:443–473.
- Zak, P. J. and Knack, S. (2001). Trust and growth. *The Economic Journal*, 111(470):295–321.

## **A Curriculum Vitae**

# Camille HÉMET

Paris School of Economics  
Office R5-29  
48 Boulevard Jourdan  
75014 Paris

camille.hemet@psemail.eu  
+33(0)1 80 52 13 55  
<http://sites.google.com/site/camillehemet>  
Update: October 2019

## Research fields

---

Urban Economics, Labor and Population Economics, Public Economics, Crime Economics

## Current position

---

**Assistant professor (*Maître de conférences*), École Normale Supérieure (Paris), since 2015**  
PSE research associate, *Labour and Public Economics* group  
CEPR research affiliate, *International Trade and Regional Economics*  
IEB associated researcher, *Cities* research program  
IPP and CEPREMAP affiliated researcher

## Former positions

---

2014 - 2015	Post-Doc, Barcelona Institute of Economics (IEB)
2012 - 2013	Pre-Doctoral research fellowship, Aix-Marseille School of Economics (GREQAM)
2009 - 2012	PhD fellowship with teaching duty, Sciences Po
Oct-12 2011	Visiting PhD student, Stanford University

## Education

---

2014	<i>Qualification</i> for assistant professorship in the French academic system
2009 - 2013	PhD in Economics, Department of Economics, Sciences Po ( <i>with highest honors</i> ) Thesis title: <i>Birds of a feather cannot always flock together. Essays on the socio-economic impacts of local diversity.</i> Jury: Yann Algan (supervisor), Bruno Deceuse (referee), Alan Manning, Thierry Mayer and Yves Zenou (referee)
2007 - 2009	Master in Economics: Economic Analysis and Policy, Paris School of Economics
2005 - 2007	Bachelor's degree in Economics and Mathematics, University of Cergy Pontoise

## Publications

---

### Peer-reviewed journals

"Diversity and Employment Prospects: Neighbors Matter!", with C. Malgouyres, *Journal of Human Resources*, Vol. 53(3), 2018, pp. 825-858.

"Next Train to the Polycentric City: The Effect of Railroads on Subcenter Formation", with M-À. Garcia-López and E. Viladecans-Marsal, *Regional Science and Urban Economics*, Vol. 67, 2017, pp. 50-63.

"How Does Transportation Shape Intrametropolitan Growth? An Answer From the Regional Express Rail", with M-À. Garcia-López and E. Viladecans-Marsal, *Journal of Regional Science*, Vol. 57(5), 2017, pp. 703-905.

"The Social Effects of Local Diversity: a Natural Experiment with Exogenous Residential Allocation", with Y. Algan and D. Laitin, *Journal of Political Economy*, Vol. 124(3), 2016, pp.696-733.

### Book chapters

"Les effets du RER sur la localisation et la croissance de l'emploi : une analyse économétrique", with M-À Garcia-López and E. Viladecans-Marsal, chapter 11 in *Les effets économiques et urbains du Grand Paris Express*, Economica, 2019.

"Diversity and Well-being: Local Effects and Causal Approaches", chapter 2 in *Social Cohesion in Diverse Societies: Mechanisms, Conditions and Causality*, Routledge, 2015.

## Works in progress

---

### Working papers

"Custodial vs. Non-Custodial Punishments: Measuring the Impact of an Anticipated Reform", with B. Michel

"Unemployment and Crime Victimization: A Local Approach"

"Ethnic Networks and the Informal Labor Market"

### Other ongoing projects

"The Strength and Use of Local Referral Networks: Evidence from France", with Clément Malgouyres

"Neighborhood Quality and Crime: Evidence from a Decade of French Urban Renewal Program", with Nina Guyon and Arnaud Philippe

"The Role of Peer Effects in Educational Choice", with Manon Garrouste

"Ethnic Diversity and Educational Attainment: an Experimental Investigation of the Underlying Mechanisms", with Arnaud Chevalier and Matthijs Oosterveen

"Local Economic Conditions and Resilience to Labour Market Shocks", with Paul Brandily-Snyers et Clément Malgouyres

## Research contracts and academic distinctions

---

### Main research contracts

- |             |  |
|-------------|--|
| 2018 - 2021 | Principal investigator and coordinator of the project <i>Social cohesion and the city: Urban infrastructure and amenities for a sustainable society</i> , funded by the French National Research Agency (ANR-18-CE22-0013-01; 165,000 €) |
| 2016 - 2019 | <i>Housing and salary: a joint analysis of two key dimensions of immigrant and refugee assimilation</i> , French National Research Agency (ANR-16-FASI-0001-01; 199,800 €), coordinated by Gregory Verdugo (CES, Paris I University)     |
| 2014 - 2016 | <i>Global challenges, local responses: the role of cities as engine of economic prosperity</i> , Ministerio de Economía y Competitividad (ECO-2013-41310-R; 20,300 €), coordinated by E. Viladecans-Marsal (UB, IEB)                     |
| 2014 - 2016 | <i>Infrastructures' Improvement and Urban Structure: the Case of Paris</i> , Société du Grand Paris (CRC-2013-06; 34,000 €), coordinated by E. Viladecans-Marsal (UB, IEB)   |

### Internal funding

- |             |   |
|-------------|---|
| 2019 - 2020 | <i>Ethnic Diversity Mechanisms</i> , PGSE (1,700 €)   |
| 2017 - 2018 | <i>Urban Renewal and Crime</i> , CEPREMAP (5,000 €)   |
| 2015 - 2016 | <i>Neighborhood diversity and income inequalities across neighborhoods</i> , Paris School of Economics, Labex OSE (2,500 €) |

### Distinctions

- |             |  |
|-------------|--|
| 2019 - 2023 | Bonus for recognition of research and PhD supervision granted by the French National University Council (PEDR) |
| 2014        | Special Distinction ( <i>Accessit</i> ) from the French Economic Association for best PhD thesis               |

## Presentation of research

---

### Invited seminars and workshops (including expected)

- 2020 Seminar: University of Lille
- 2018 Seminars: University of Groningen, IZA (Bonn), BETA (Nancy), SERC-LSE (London), ULB (Charleroi), University of Bern, Fourgeaud seminar (French Ministry of Economy and Finance)
- 2017 Seminars: SSMSI (French Ministry of Interior), RUES-PSE (*discussant*), French Central Bank (*discussant*)
- 2016 Seminars: CIRED (Paris), BETA (Strasbourg), GATE (Saint-Etienne), University of Lund, Ministry of environment, RUES-PSE (*discussant*)  
Workshop on Cultural Diversity (Paris 1), IV Workshop on Urban Economics (*discussant*), Migration workshop (PSE*discussant*),
- 2015 Seminars: CRED (Paris 2), LEDa (Paris 9), LED (Paris 8), PSE, LEMMA (Paris 2)  
Summer school "Social Interactions and Urban Segregation" (CREM, Univ. Rennes)
- 2014 Seminars: GATE (Lyon), CBS (Copenhague), Uppsala University, IEB (Barcelona)  
III Workshop on Urban Economics (IEB, Barcelona)
- 2013 Seminars: NUS (Singapore), PSE (Paris), Sciences Po (Paris)  
Workshop on Ethnic Diversity and Social Capital (WZB, Berlin)
- 2012 Seminar: AMSE-GREQAM (Marseille)  
II Workshop on Urban Economics (IEB, Barcelona), Workshop on Immigration, Ethnicity and Urban Inequality in Europe (UCLA)
- 2010 Seminar: Sciences Po  
Symposium on Ethnic Diversity and Public Goods (Stanford), Workshop on Discrimination and Segregation (Sciences Po)

### Conferences

- 2018 Multidisciplinary conference on "African Cities", organized at ENS. Talk on urban economics principles.
- 2015 UEA (Portland), ERSa (Lisbon), LAGV days (Aix en P.), ADRES doctoral meeting (Paris)
- 2014 UEA (Washington), AFSE (Lyon), SOLE (Arlington), ADRES doctoral conference (Paris)
- 2013 AFSE (Aix en Provence), LAGV days (Aix en Provence), AMSE PhD workshop (Aix), NORFACE Migration Conference (Londres), ADRES doctoral conference (Strasbourg)
- 2012 EEA (Malaga), ESEM (Malaga), AFSE (Paris), ESPE (Berne), ADRES (Marseille)
- 2011 International Conference on Shadow Economy (Münster)

## Non-academic dissemination of research

---

### Presentations of my research on transportation, local employment growth and decentralization:

- French Ministry of Finance: Fourgeaud seminar on "Transportation and urban shape", January 2018
- *Printemps de l'économie*: round table on "The effects of local public policies", April 2016
- Sustainable Development division of the French Ministry of Environment: 21<sup>st</sup> CGDD-PSE seminar devoted to "Transportation, jobs location and agglomeration economies", April 2016

### Presentations of my research on the local determinants of crime victimization:

French Ministry of Interior, November 2017

### Short notes (in French) generalizing research results for a large audience:

- "Les infrastructures de transport sont-elles un vecteur de croissance des villes ?", in *5 articles...en 5 minutes !* (monthly series of 5 synthesis of recent articles by PSE researchers published on PSE website), November 2017
- "Comprendre les effets de la diversité des origines sur les conditions de logement via les relations sociales entre voisins : le cas des quartiers HLM en France", in *5 articles...en 5 minutes !*, February 2016
- "L'approche békérienne de la transgression des lois", in *Regards Croisés sur l'Économie*, March 2014
- "L'immigration et le travail non déclaré", in *Regards Croisés sur l'Économie*, March 2014

## Refereeing activities

---

### Academic journals

Annals of Economics and Statistics, Economics Bulletin, European Economic Review, Journal of Ethnic and Migration Studies, Journal of Housing Economics, Journal of Human Resources, Journal of Public Economics, Journal of Public Economic Theory, Journal of Regional Science, Journal of Urban Economics, Labour Economics, Regional Science and Urban Economics, Review of Economics and Statistics, Revue Économique, Socio-Economic Review

### Other expertise:

National Research Foundation of Luxemburg

## Collective responsibilities

---

Since 2019	Elected member of the scientific council of PSE research lab (UMR PjSE)
2019	Member of Scientific committee: 9th European Meeting of the UEA
2018	Member of Scientific committees: 8th European Meeting of the UEA and 67 <sup>th</sup> meeting of the French Economic Association
Since 2017	Member of hiring committees for junior positions ( <i>Maîtres de Conférences</i> ) in French universities: Université de Grenoble (2017), École Normale Supérieure de Lyon (2018), Université Paris 1 Panthéon-Sorbonne (2018), Université Paris-Est Créteil (2018), Université Aix-Marseille (2018, 2019)
	Member of the consultative committee for the French Victimization Survey
Since 2017	Head of Studies of ENS Economics Department
2017 - 2019	Organizer of the selection committee for ENS students ( <i>concours voie universitaire</i> )
Since 2016	Co-organizer of the PSE Labour and Public Economics external and internal seminars
Since 2015	Participation in the training of PSE job market candidates (mock interviews) Co-organizer of a seminar around economy-related news for ENS students
2014 - 2015	Organization of IEB external seminars Presentation on "How to prepare to the job market?" for UB PhD students
2014	Member of the organization team of the III Workshop on Urban Economics, IEB
2012 - 2013	Organizer of an informal seminar for AMSE PhD students and researchers

## Teaching

---

### Lectures

Since 2019	Urban Economics (shared with Laurent Gobillon), Master Economic Decision and Cost-Benefit Analysis, PSE Urban Economics, 3-hours doctoral course, PSE
2017 - 2019	Population dynamics (3/5 of a course shared with Gilles Pison and Laurent Toulemon), Master Public Policies and Actions for Sustainable Development, ENPC
Since 2016	Public Economics (intermediate), Master in Public Affairs, Sciences Po
Since 2015	Advanced microeconomics, ENS, L3 (except in 2016) Introduction to economics, ENS (for non-economists)
2015 - 2017	Introduction to research in economics, ENS, L3
2014 - 2015	Spanish Economics, University of Barcelona, L3 - <i>in English</i>
2013 - 2015	Local public finance, University of Barcelona, L3 - <i>in English</i>
2010 - 2011	Static and dynamic optimization, Sciences Po, M1 (EPP)
2009 - 2010	Introduction to mathematical and quantitative methods, Sciences Po, L1

### Practice classes and other teaching activities

- 2011 - 2012 Introduction to macroeconomics (TA), Sciences Po, L1  
2010 - 2011 Introduction to macroeconomics (Practice class and TA), Sciences Po, L1  
Public economics (TA and practice), Sciences Po, L2 - *in English*  
Introduction to microeconomics (Practice class and tutoring), Sciences Po, L1  
2009 - 2010 Introduction to macroeconomics (Practice class, tutoring, on-line courses), ScPo, L1  
2008 - 2009 International trade (Practice class), Paris I University, L3

### **Students supervision and related activities**

---

#### Member of PhD defense jury

- 2020 Mariona Segú (Univ. Paris-Saclay), supervised by Miren Lafourcade (Univ. Paris-Saclay) and Gabrielle Fack (Univ. Paris-Dauphine)  
2019 Gudrun Johnsen (ENS), supervised by Gilles Saint-Paul (ENS, PSE)  
2018 Andrea Aguirre Sanchez (University of Saint-Etienne):  
*Urban Crime in Ecuador: Three essays on the role of economic inequalities, population density and emotions,*  
Supervised by Nelly Exbrayat and Stéphane Riou  
2015 Simón Planells Struse (Univ. of Barcelona):  
*Essays on the Economics of Crime: Determinants of Crime in an Urban Context,*  
Supervised by D. Montolio Estivill (IEB)

#### Member of PhD dissertation committees

- 2019 Mariona Segú, supervised by Miren Lafourcade (Univ. Paris-Saclay) and Gabrielle Fack (Univ. Paris-Dauphine)  
Gudrun Johnsen, supervised by Gilles Saint-Paul (ENS, PSE)  
2018 Benjamin Carantino, supervised by Miren Lafourcade (U. Paris-Saclay)  
2017 Andrea Aguirre Sanchez, Supervised by Nelly Exbrayat (U. Saint-Etienne)

#### Supervision of Master's students

- 2017-2018 Madeleine Péron, Master's dissertation (APE program, PSE): *Does Unemployment Cause Crime?*  
2016-2017 Marie Gaudier, Master's dissertation (APE program, PSE): *Is Crime an Environmental Phenomenon? The Role of Local Equipments on Crime Occurrence in French Municipalities.*  
Since 2018 Co-supervision (with François Fontaine) of a research seminar for PSE Master's students

#### Other

- Since 2016 Individual advisor (tutor) for ENS students

## **B List of Published and Unpublished Works**

### **Articles published in peer-reviewed journals**

"Diversity and Employment Prospects: Neighbors Matter!", with C. Malgouyres, *Journal of Human Resources*, Vol. 53(3), 2018, pp. 825-858.

"Next Train to the Polycentric City: The Effect of Railroads on Subcenter Formation", with M-À. Garcia-López and E. Viladecans-Marsal, *Regional Science and Urban Economics*, Vol. 67, 2017, pp. 50-63.

"How Does Transportation Shape Intrametropolitan Growth? An Answer From the Regional Express Rail", with M-À. Garcia-López and E. Viladecans-Marsal, *Journal of Regional Science*, Vol. 57(5), 2017, pp. 703-905.

"The Social Effects of Local Diversity: a Natural Experiment with Exogenous Residential Allocation", with Y. Algan and D. Laitin, *Journal of Political Economy*, Vol. 124(3), 2016, pp.696-733.

### **Working papers**

"Custodial vs. Non-Custodial Punishments: Measuring the Impact of an Anticipated Reform", with Bastien Michel. Latest version: October 2019.

"The Strength and Use of Local Referral Networks: Evidence from France", with Clément Malgouyres. Latest version: March 2019.

"Unemployment and Crime Victimization: A Local Approach". Latest version: November 2018.

"Ethnic Networks and the Informal Labor Market". Latest version: June 2017.

### **Other ongoing projects**

"Neighborhood Quality and Crime: Evidence from a Decade of French Urban Renewal Program", with Nina Guyon and Arnaud Philippe

"The Role of Peer Effects in Educational Choice", with Manon Garrouste

"Ethnic Diversity and Educational Attainment: an Experimental Investigation of the Underlying Mechanisms", with Arnaud Chevalier and Matthijs Oosterveen

"Local Economic Conditions and Resilience to Labour Market Shocks", with Paul Brandily-Snyers et Clément Malgouyres

### **Book chapters**

"Les effets du RER sur la localisation et la croissance de l'emploi : une analyse économétrique", with M-À Garcia-López and E. Viladecans-Marsal, chapter 11 in *Les effets économiques et urbains du Grand Paris Express*, Economica, 2019.

"Diversity and Well-being: Local Effects and Causal Approaches", chapter 2 in *Social Cohesion in Diverse Societies: Mechanisms, Conditions and Causality*, Routledge, 2015.

## C Presentation of research

### External seminars

Applied Economics lunch seminar, Paris School of Economics - PSE (2013); National University of Singapore (2013); Groupe d'Analyse et de Théorie Economique - GATE, Lyon (2014); Copenhagen Business School (2014); University of Uppsala University (2014); Centre de Recherches en Economie du Droit - CRED, Paris 2 University (2015); Laboratoire d'Economie de Dauphine - LEDa, Paris 9 University (2015); Laboratoire d'Economie Dionysien - LED, Paris 8 University (2015); PSE (2015); Laboratoire d'Économie Mathématique et de Microéconomie Appliquée - LEMMA, Paris 2 University (2015); Centre International de Recherche sur l'Environnement et le Développement - CIRED, Nogent-sur-Marne (2016); Bureau d'Economie Théorique et Appliquée - BETA, Strasbourg (2016); GATE, Saint-Etienne (2016); University of Lund (2016); University of Groningen (2018); Institute of Labor Economics - IZA, Bonn (2018); BETA, Nancy (2018); Spatial Economics Research Center, London School of Economics - SERC, LSE (2018); Université Libre de Bruxelles - ULB, Charleroi (2018); University of Bern (2018); University of Lille (scheduled in 2020).

### Workshops

Symposium on Ethnic Diversity and Public Goods, Stanford University (2010); Workshop on Discrimination and Segregation, Sciences Po (2010); Workshop on Immigration, Ethnicity and Urban Inequality in Europe, University of California in Los Angeles (2012); II Workshop on Urban Economics, IEB, Barcelona (2012); Workshop on Ethnic Diversity and Social Capital, WZB, Berlin (2013); III Workshop on Urban Economics, IEB, Barcelona (2014); Workshop on Social Interactions and Urban Segregation, Rennes (2015); Workshop on Cultural Diversity, Paris 1 University (2016).

### Conferences

International Conference on Shadow Economy, Münster (2011); European Economics Association - EEA, Malaga (2012); European meeting of the Econometric Society, ESEM, Malaga (2012); Association Française de Sciences Economiques - AFSE, Paris (2012); European Society of Population Economics - ESPE, Berne (2012), Association pour le Développement de la Recherche en Économie et Statistiques - ADRES, Marseille (2012); AFSE, Aix en Provence (2013); Louis-André Gérard Varet - LAGV days, Aix en Provence (2013); NORFACE Migration Conference, London (2013); ADRES, Strasbourg (2013); Urban Economics Association - UEA, Washington (2014); AFSE, Lyon (2014); Society of Labor Economics - SOLE, Arlington (2014); ADRES, Paris (2014); UEA, Portland (2015); European Regional Science Association - ERSA, Lisbon (2015); LAGV days, Aix en Provence (2015); ADRES, Paris (2015); Multidisciplinary conference on "African Cities", ENS, Paris (2018).

### Non-academic presentations of research

Round table on "The effects of local public policies" during the *Printemps de l'économie* (2016); 21<sup>st</sup> CGDD-PSE seminar devoted to "Transportation, jobs location and agglomeration economies", Sustainable Development division of the French Ministry of Environment (2016); Statistical Unit of the French Ministry of Interior - SSMSI (2017); Fourgeaud seminar on "Transportation and urban shape", French Ministry of Economy (2018).

## **D Research Articles**

This section is a collection of the research articles presented in the above dissertation. It consists in four papers published in peer-reviewed journals, and four working papers. The papers are presented in the order in which they appear in the dissertation.

**D.1 The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation**

# The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation

---

Yann Algan

*Sciences Po*

Camille Hémet

*Institute of Economics of Barcelona*

David D. Laitin

*Stanford University*

Relying on diversity measures computed at the apartment block level under conditions of exogenous allocation of public housing in France, this paper identifies the effects of ethnic diversity on social relationships and housing quality. Housing Survey data reveal that diversity induces social anomie. Through the channel of anomie, diversity accounts for the inability of residents to sanction others for vandalism and to act collectively to demand proper building maintenance. However, anomie also lowers opportunities for violent confrontations, which are not related to diversity.

## I. Introduction

Recent research has drawn an ominous picture of the implications of cultural heterogeneity on social peace and economic growth. A large lit-

The research leading to these results has received funding from the European Research Council for Yann Algan under the European Community's Seventh Framework Program

Electronically published May 4, 2016  
[*Journal of Political Economy*, 2016, vol. 124, no. 3]  
© 2016 by The University of Chicago. All rights reserved. 0022-3808/2016/12403-0003\$10.00

erature shows a negative relationship, though not always robust, between ethnic diversity and the quality of public goods (Alesina, Baqur, and Easterly 1999; Alesina and La Ferrara 2000; Miguel 2004; Miguel and Gugerty 2005), welfare spending (Luttmer 2001), civil conflict and trust (Fearon and Laitin 2000; Alesina and La Ferrara 2002; Putnam 2007), and economic growth (Alesina et al. 1999). The leading explanations of why ethnic fragmentation affects those outcomes are the heterogeneity of preferences and the free-rider problem, which undermines collective action. The literature thus views the problem of fractionalization in terms of voting behavior on aggregate outcomes such as public goods provision at the country or county level. Yet, little is known on how diversity directly affects social relationships and well-being at the neighborhood level. Our paper fills this gap by looking at the effect of ethnic diversity on the quality of common spaces through social relations within local communities at the housing block level. Moreover, we provide a new identification strategy to overcome the endogeneity problem raised by residential self-selection. We rely on a natural experiment of exogenous spatial allocation in the French public housing sector to identify the causal effect of diversity on those outcomes.

The main contribution of our paper is to identify the effect of ethnic diversity on social relationships and the quality of public goods at a very local block level. We use microdata on housing conditions in which the units of observation are public housing blocks, defined as sets of houses or apartment buildings delimited by the surrounding streets.<sup>1</sup> This is a key improvement for the analysis of how diversity shapes social relationships compared to the previous literature that is based on aggregated data at the county, regional, or country level. Diversity might matter for various reasons at different levels, and the channels through which diversity operates are likely to depend on the size of the unit of observation. By focusing on the provision of public goods at an aggregate level, the pre-

---

(FP7/2007–2013)/ERC grant agreement 240923. The authors acknowledge financial support from the France-Stanford Center for enabling this cross-Atlantic collaboration. We are grateful to Alberto Alesina, Ernesto Dal Bo, Larry Katz, Asim Ijaz Khwaja, Edward Miguel, Leah Platt Boustan, Thierry Mayer, Daniel Posner, Andrei Shleifer, Mathias Thoenig, and Yves Zenou for valuable comments and discussions. We also thank seminar and conference participants at the National Bureau of Economic Research Political Economy seminar, at Stanford University, at Sciences Po, at the French Research Center in Economics and Statistics, at the French Association for Developing Research in Economics and Statistics, at the University of California, Los Angeles, at the Institute of Economics of Barcelona, at the European Economic Association, at Aix-Marseille School of Economics, and at Uppsala University for useful comments. We are grateful to Corinne Prost, former head of the Employment Division at the French National Statistical Institute (INSEE) for allowing access to key data through a convention between INSEE and Sciences Po. Data are provided as supplementary material online.

<sup>1</sup> Our units of analysis, called *ilots* in French, are in fact defined similarly to US census blocks.

vious literature is mainly interested in the effect of diversity on collective action through lobbying or patronage (see Alesina and La Ferrara [2005] for a survey). Instead, we analyze in this paper how diversity within a small community affects individual well-being and satisfaction with housing conditions through relationships among neighbors. We exploit the French Housing Survey, which reports specific information about the neglect and voluntary degradations of the public areas, the quality of the housing, and interpersonal conflicts between neighbors. These data make it possible for the first time to identify various effects of diversity on local social relationships and public goods outcomes and to explore the possible channels explaining this link.

When residents of more diverse blocks report that neglect and voluntary degradations are rife in their housing unit, we interpret this as a result of the residents' failure to develop social norms and other-regarding preferences. When they report the breakdown and the poor quality of basic facilities (such as heating and soundproofing), we interpret this as a result of a diminished capacity for collective action for social improvement. Those goods are of course not directly degraded by diversity. But diversity might be associated with lower ability for collective action, explaining the irregularity of maintenance and the absence of repairs in more diverse blocks. In this case, the result could be supported in equilibrium if the landlords reckon that they can neglect facilities in ethnically heterogeneous housing projects, knowing that they will not face collective action from their tenants demanding better services. Finally, when residents report incidents of direct interpersonal conflicts, we can interpret this as an effect of diversity on cultural enmity. We test these channels by using indirect and objective measures of the quality of social relationships and common spaces, such as the number of repairs and the upkeep of the security equipment. We also exploit data on municipal police and show that local police resources are higher in more diverse areas. We find that diversity decreases the quality of local common spaces but has no effect on public safety. We also show that individuals are more likely to report the absence of any social relationship with their neighbors than interpersonal conflicts with them. We thus conclude that diversity leads to social anomie, preventing the emergence of social norms and collective action.

In order to make unbiased causal inferences, we provide a new strategy for identifying the causal effect of diversity on economic and social outcomes. The general concern in this literature is that the endogenous residential sorting of individuals on ethnic grounds biases the estimate of the impact of diversity. We address this issue by using a natural experiment in which households in France are allocated to public housing blocks without taking their ethnic origin or their preference for diversity into account. Owing to a strongly republican ideology, the French public

housing system allocates state-planned moderate-cost rental apartments (HLMs—*habitations à loyer modéré*) to natives and immigrants without concern for their cultural and ethnic background, mixing people indiscriminately. Some HLM neighborhoods are consequently quite diverse and others quite homogeneous. Furthermore, HLM inhabitants rarely move, as the rents are much lower than market rates. Consequently, residents cannot choose whether to live near people like themselves. Rather, they accept their placement, whether next to coethnics or strangers. Methodologically, this means that we can take the degree of diversity in any one HLM block as exogenous, connect the level of diversity with the housing situation, and examine whether greater heterogeneity leads to poorer provision of public goods or more troubled social relationships in French communities. We extensively document the actual process of allocation of households within the public housing sector. We show that legal rules prohibit housing allocation based on ethnic backgrounds and that, in practice, the characteristics of the public housing sector make it very complicated to bypass the law.

We also conduct a variety of formal statistical tests to verify the absence of self-sorting on ethnic characteristics. In particular, we run various placebo tests at the housing block level on housing characteristics that logically cannot be related to diversity, that is, fixed characteristics over which residents cannot have any control. We show that diversity does not correlate with measures of exogenous characteristics of the distribution of public housing characteristics. We perform a variety of alternative tests. Focusing on households that moved into a public housing unit in the previous year, we do not find any evidence of self-segregation along ethnic lines. We also examine potential self-selection prior to the move and show that households that have refused an offer end up living in public housing blocks that display the same ethnic diversity as those that accepted their first offer. Thus even if some households were willing to be choosy with respect to the ethnic composition of their neighborhoods, they cannot self-segregate in the public housing sector because of the allocation process and the tight supply constraints of dwellings.

Naturally, this paper is not the first one to try to overcome this identification issue. But other attempts to establish causality rely mainly on instrumental variables.<sup>2</sup> However convincing the instruments might be, this strategy cannot overcome the concern as to whether the instruments fulfill the exclusion restriction and do not have a direct effect on public goods. For instance, Miguel (2004) and Miguel and Gugerty (2005) use

<sup>2</sup> In their seminal contribution to the literature, Alesina et al. (1999) provide a first attempt to deal with this endogeneity issue by collecting data at different levels of aggregation (cities, metropolitan areas, and counties). Their assumption is that different levels of aggregation allow for the correction of the potential biases introduced by Tiebout sorting.

the precolonial patterns of settlement as instruments, assuming that these variables have no direct impact on present-day ethnic relations. More recently, Glennerster, Miguel, and Rothenberg (2013) have also relied on historical data of fractionalization as an instrument. But since past settlement patterns are likely to have at least some direct impact on present-day ethnic relations, the exclusion restriction might still be technically violated. Using a natural experiment with exogenous allocation of ethnic groups is thus an alternative strategy to deal with these traditional caveats. The paper that is the closest to ours is Dahlberg, Edmark, and Lundqvist (2012), which uses a nationwide policy intervention program that exogenously placed refugees coming to Sweden across Swedish municipalities. However, their paper examines in-group bias in preferences for redistribution rather than the effect of diversity on local public goods and social relationships.

Our paper is related to the large literature on the effects of ethnic diversity on economic and social outcomes. In US cities, higher ethnic diversity has been found to be associated with lower social capital (Alesina and La Ferrara 2000, 2002; Putnam 2007), lower welfare spending (Luttmer 2001), and poorer quality of public goods (Alesina et al. 1999). In western Kenya, the greater the mixing of tribes, the less people have public spiritedness, and the lower the contributions to public goods (Miguel 2004; Miguel and Gugerty, 2005). In cross-national surveys, diversity correlates with low growth in GDP and low quality of institutions (Easterly and Levine 1997; Alesina et al. 2003). Alesina and Zhuravskaya (2011) show that islands of homogeneity amid a broadly diverse country do not decrease the negative effects of diversity on the quality of government.<sup>3</sup> Theoretical contributions, in particular on ethnic conflicts, can be found in Esteban and Ray (2011) and Caselli and Coleman (2013). These findings are depressing, in a normative sense, for those who herald gains from diversity (Page 2007) and depressing, in an empirical sense, as in our globalized world, local cultural diversity is increasingly common (Dancygier 2010). However, the robustness of the relationship and the channels at work remain to be determined. Putnam (2007) is careful to underline that his data allow him to claim only short-run correlation between diversity and

<sup>3</sup> The magnitude of the relationship between those outcomes and ethnic diversity is substantial. Putnam (2007) finds that the difference between living in a highly homogeneous city (Bismarck, North Dakota) and heterogeneous Los Angeles is as great as the difference between an area with a poverty rate of 7 percent and one with a poverty rate of 23 percent. Alesina et al. (1999) show that moving from complete homogeneity to complete heterogeneity is associated with a reduction in spending on roads by 9 percentage points. Luttmer (2001) finds that interpersonal preferences based on negative exposure and racial group loyalty of recipients are associated with 33 percent of the cross-state variation in the support for welfare spending. Alesina et al. (2003) show that moving from perfect homogeneity to maximum heterogeneity would be associated with a reduction in a country's growth rate by 2 percentage points per year.

trust. Miguel (2004) finds no diversity impacts on local outcomes in Tanzania, a country in which the ruling authorities have sought to ameliorate ethnic cleavages by promoting a common language. Posner (2004) shows that changed electoral rules can create broader ethnic identities, thereby reducing fragmentation. Dunning and Harrison (2010) show that intertribal polarization in Mali is reduced with cross-cutting cleavages. Glennerster et al. (2013) also argue that the presence of strong chiefs at the local level, although reinforcing the salience of ethnicity, translates into effective interethnic cooperation.<sup>4</sup> Finally, Alesina and La Ferrara (2005) document the potential positive effect of diversity on productivity through complementarity in skills.

Our paper is also incidentally related to empirical works examining neighborhood effects on social and economic outcomes. So far, the literature has mainly focused on the neighborhood effects on physical and mental health, economic self-sufficiency, risky and criminal behavior, or educational outcomes (see, among many others, Katz, Kling, and Liebman [2001], Oreopoulos [2003], Goux and Maurin [2007], and Kling, Liebman, and Katz [2007]). In particular, Katz et al. (2001) and subsequent contributions use the Moving to Opportunity program to estimate the externalities from neighbors. To avoid the problem of endogenous neighborhood selection, those authors use data from this randomized experiment in which some families living in high-poverty US housing projects were offered housing vouchers to move to higher-income areas. While our paper is not based on a randomized experiment, we also avoid the inferential issues of residential endogenous selection by using the exogenous spatial allocation of households with respect to ethnic characteristics. We enlarge the dimensions analyzed in this literature by looking at how immediate neighborhood diversity affects well-being and the quality of the local environment.

The remainder of this paper is organized as follows. Section II describes the data. Section III presents our identifying assumption on the absence of residential self-sorting in public housing: we document the allocation process in the French public housing market, providing institutional support for our assumption. We then demonstrate that at the block level, diversity does not correlate with measures of exogenous characteristics of the distribution of public housing characteristics. Section IV shows our main results. We document the effects of ethnic diversity on residents' satisfaction with their housing conditions, local public goods quality, and social relationships. We discuss the various dimensions and channels through which diversity might matter for households' well-being at the local level

<sup>4</sup> Varshney (2003) and Jha (2013) also show how local institutions can ameliorate communal violence in India.

in Section V. Section VI provides robustness tests on the validity of self-reported outcomes. Section VII presents conclusions.

## II. Presentation of the Data

### A. Data Sets

We rely on two representative French national surveys to estimate the relationship between ethnic diversity and the quality of public space within the housing block.<sup>5</sup> In each survey, we focus on the subsample of public housing population in order to identify the causal effect of diversity and control for self-sorting. Our main data source is the French housing survey of 2002 (Enquête Logement, from the French Statistical Institute [INSEE]; hereafter HS), which provides detailed information on the intensity and quality of social relationships with neighbors and on the quality of local public spaces, ranging from vandalism in the common areas to housing quality and conflicts in the neighborhood. The HS also reports detailed information about the ethnic, economic, and social backgrounds of surveyed households.<sup>6</sup> The 2002 wave of the HS surveys more than 32,000 households, out of which about 16 percent live in the public housing sector. Our sample thus contains 5,189 observations (households) living in about 2,500 different blocks. The sampling of the HS is such that not all the individuals living in a given block are systematically surveyed, but are randomly drawn instead. Therefore, we cannot compute any representative measure of block-level characteristics (in particular, diversity) using the HS data. We overcome this concern by using a second database, the 1999 French Population Census, which is an exhaustive survey covering the entire population living in France. Each HS sample is drawn from the most recent census, and the geographical units of the HS are a subsample of those of the census. There are, on average, 2,895 blocks with public housing tenants per *département* in the census (the median number is 741). The mean (respectively, median) number of public housing tenants in these blocks is 18.4 (respectively, eight). The census provides variables such as birth country or nationality at birth, from which we compute a representative measure of diversity at the housing block level, which we are then able to match with the corresponding housing block in the HS. The census also provides some information about building characteristics that will be used in Section III.B to test our identifying assumption.

<sup>5</sup> A third survey, the Labor Force Survey, is used to perform some of the tests presented in the online appendix and hence is described in this appendix.

<sup>6</sup> Some of the key variables for our study are not public. INSEE made their access possible as part of a convention between the INSEE and Sciences Po. We were required to make use of the “sensitive” data within the confines of the INSEE.

*B. Fractionalization Indexes and Sample Characteristics*

We measure ethnic diversity with the standard fractionalization index (hereafter DIV) used in the literature (see, e.g., Alesina et al. [2003] for a detailed description).<sup>7</sup> This index reflects the probability that two randomly drawn individuals from a given population belong to different groups (previous studies looked at ethnolinguistic or religious groups, while we focus on diversity in terms of national origins). More formally, the basic fractionalization index is computed as one minus the Herfindahl index of group shares:<sup>8</sup>

$$\text{DIV}_l = \sum_{i=1}^{i=N} s_{il}(1 - s_{il}) = 1 - \sum_{i=1}^{i=N} s_{il}^2, \quad (1)$$

where  $s_{il}$  is the share of group  $i$  ( $i = 1, \dots, N$ ) in area  $l$ . If the population living in area  $l$  is fully homogeneous,  $\text{DIV}_l$  equals zero, and it converges to one as the population heterogeneity increases. Note that  $\text{DIV}_l$  can increase for two reasons: it will increase with the number of groups, and it will increase the more equal the size of the groups. As mentioned above, the census provides information about the country of birth and the nationality at birth of individuals, allowing us to construct two different measures of diversity. In the remainder of the paper, we focus on diversity by nationality at birth, computed at the block level. The distribution of diversity faced by tenants living in public housing blocks is presented in Appendix table A1. The average public housing tenant surveyed in the 1999 census lives in a block with 28 percent diversity. After matching this measure of diversity to the corresponding public housing blocks in the HS, we obtain that the average public housing tenant lives in a block with 25 percent diversity. The highest level of diversity observed in a public housing block is 87.5 percent in the census and 80.2 percent in the HS.

Appendix table A2 presents the main sociodemographic characteristics of the public housing sample from the 2002 HS. Foreigners (or immigrants) are overrepresented in the public housing population compared to the private housing population. Public housing neighborhoods are also characterized by a poor socioeconomic environment, where individuals have low education levels and earn low incomes: around one-third of adults have no diploma at all, and the share of individuals having achieved graduate studies is 12 percent, less than half the corresponding share in the private housing sector. The bottom of table A2 also reports some characteristics of the living environment of the surveyed public

<sup>7</sup> We have also tried alternative indices such as polarization indicators, yielding similar results.

<sup>8</sup> These groups can be defined by, among other things, ethnicity, language, nationality, or country of origin.

housing families: slightly more than half of households live in buildings built between 1949 and 1974, and the average household lives in a block where there is a 22 percent unemployment rate, more than twice as large as the national average in 1999, the year in which the block unemployment rate was computed.

### III. The Exogeneity of Diversity in the Public Housing Sector

This section addresses the main identification issue raised by the estimation of the causal impact of ethnic diversity on social interactions and the quality of public goods. The issue, common to all the literature on ethnic diversity, is that fractionalization presents a high risk of endogeneity. Individuals generally tend to self-segregate: they prefer forming links with others like themselves, with whom they share common interests, and in particular people of the same ethnicity or the same social background.<sup>9</sup> If people can choose the area where they live, diversity would be an outcome of strategic choices, and attempts at measuring the effects of diversity would be confounded. If all people would rather move into neighborhoods where people are similar to themselves and richer people could better afford to move, we would observe a (spurious) relationship of diversity and wealth. But if wealthy families that live in diverse settings are those that have a taste for diversity, the true effect of diversity on social outcomes would be an underestimate. Therefore, the level of diversity of the neighborhoods is probably endogenous, and any estimates on the implications of diversity will be biased.<sup>10</sup>

To identify the effect of ethnic diversity, one must therefore study individuals who are assigned to their place of residence without consideration of ethnic characteristics. The purpose of this section is to bring forth evidence that spatial allocation of households across public housing blocks in France can be considered as exogenous with respect to ethnic characteristics due to French regulation. Naturally, the sample of households that apply for public housing dwellings is endogenous with respect to economic, social, or cultural characteristics. But among the pool of selected households, we show that their spatial allocation across the public housing blocks of a given *département* is exogenous with respect to their ethnic characteristics, conditional on their other characteristics.

Note that the mere fact of working at the block level already decreases the extent of endogenous sorting. First, although households can generally choose the neighborhood to which they move, they may not be able

<sup>9</sup> Self-sorting is most typically based on race or ethnicity.

<sup>10</sup> Combes et al. (2012) use customer discrimination theory to show that landlords will tend to discriminate against ethnic minorities when renting their apartments, bringing new evidence as to why any causal claim of ethnic diversity on public goods in the private housing market would be biased.

to select a particular block in this selected neighborhood. Second, while it is possible to have an idea of the socioeconomic and demographic characteristics of a given neighborhood, it is much more difficult to observe such characteristics in a specific block prior to moving. Bayer, Ross, and Topa (2008) rely on this key assumption and provide empirical evidence that individuals' characteristics are not correlated within blocks. We present a more formal discussion and a statistical test of the exogeneity of the allocation process with respect to ethnic characteristics below.

*A. An Ethnically Blind Allocation Process Built into Law*

We first document the actual process of allocation of households across public housing dwellings. This gives a legal basis to our identifying assumption of the absence of self-sorting on ethnic characteristics in the public housing sector.<sup>11</sup> In France, the only eligibility requirements for admittance into the public housing sector are to be legally living in France (as a French citizen or migrant with a valid residence permit) and to be living under a certain threshold of income per equivalent household member.<sup>12</sup> This income ceiling is rather high: in 2009, this threshold was between €36,748 and €50,999 per year for a four-person family, depending on the region of residence (the upper figure being nearly €3,000 higher than the average disposable income of four-person households in 2007). Using the 2002 HS data, Jacquot (2007) estimates that given their income, between two-thirds and four-fifths of households living in continental France could apply for a public housing unit. As a consequence, the population eligible for public housing is about three times as large as the available space in vacant dwellings. As a result of the boom in housing prices in the private sector during the mid-1990s and the 2000s, the public housing market became even more attractive, luring new categories of people who could no longer afford to live in the private housing market. This inflow increased further the applicants to vacant units ratio. This implies that other criteria must be taken into account in the selection process. First, household size is considered to ensure a suitable match with the characteristics of vacant dwellings. More importantly, the degree of emergency of the application is taken into account. To administer this, five priority criteria are defined by law at the national level to ensure that

<sup>11</sup> The process of allocation across public housing blocks in France was mainly inspired by theories from Le Corbusier (1887–1965). Le Corbusier insisted that France must avoid the homogeneous ghettos of the urban landscapes elsewhere and should therefore allocate housing blind to ethnicity, not permitting family networks to grow within housing establishments. These ideas were translated into state regulation (Bernardot 2008).

<sup>12</sup> To compute the income per equivalent household member, the INSEE weights each household member as follows: 1 for the first adult, 0.5 for any other person of 14 years old or older, and 0.3 for any other person younger than 14 years old.

vacant housing will first be attributed to households with obvious social difficulties: those in which there is a mentally or physically disabled person, those living in precarious or hazardous shelters because of financial constraints, those living in a temporary accommodation, individuals living in a precarious shelter who recently found a job after a long unemployment spell, and spouse-abused individuals.<sup>13</sup>

To get in the queue for a housing unit, households submit a form ([https://www.formulaires.modernisation.gouv.fr/gf/cerfa\\_14069.do](https://www.formulaires.modernisation.gouv.fr/gf/cerfa_14069.do)) containing the following information: name, date of birth, family situation, employment status, resources of the household, reasons for applying to the public housing sector (currently or soon to be homeless, or reasons related to a health situation, family situation, job situation, inappropriate current housing, or unpleasant environment), type of housing looked for, whether the applicant is disabled, and whether this is the first application. It is important to stress the fact that the application form contains very limited information about the ethnicity of the applicants: they need to inform only about their nationality, which is limited to three possible categories (French, EU, or non-EU).

Entering a public housing unit results in the cancellation (*radiation*) of the application. Therefore, when public housing tenants want to move to another public housing unit, they have to go through the same whole procedure as if this were the first demand and are given a new application number. The application form is the same for everyone, independently of whether the household is already a public housing tenant or not, and simply includes a box indicating whether the current unit is in public housing or not. Since both public housing and private housing applicants go in the same pool of applicants, the same criteria apply: resources, family structure, and the five priority criteria (although the criteria of living in a precarious shelter or in a temporary accommodation are unlikely to apply for applicants already living in the public housing sector).

We now document the selection process of the applicants. The commissions in charge of allocating households to vacant public housing dwellings are held at the *département* level (or at the city level in the case of Paris, which is both a city and a *département* because of its size).<sup>14</sup> Their composition is regulated by law: a commission includes six members of the pub-

<sup>13</sup> Article L441-1 of the law relative to construction and housing, Code pour la Construction et l'Habitat.

<sup>14</sup> Continental France is divided into 22 large administrative areas, called *régions* (regions henceforth), and into 96 smaller administrative areas, called *départements*. Each *département* is hence a subdivision of a region, and several *départements* can belong to the same region. Each *département* is administered by an elected General Council (Conseil Général) and its president, whose main areas of responsibility include the management of a number of social and welfare programs, primary and secondary schools, buildings and technical staff, local roads, rural buses, and municipal infrastructure.

lic housing offices board, a representative of associations promoting integration and housing for disadvantaged people,<sup>15</sup> mayors of the municipalities in which vacant housing is to be allocated, and a representative of any association defending tenants' rights. An additional *département* representative may be part of the commission. For each vacant housing unit, at least three households must be considered by the commissioners, who finally decide which household will be allocated to the vacant housing unit considered, according to the eligibility and priority criteria detailed above. Other criteria such as the number of children in the household are also taken into account in order to allocate suitable dwellings.<sup>16</sup>

With the allocation process regulated by legal rules at the national level, it seems unlikely that households can be allocated according to their origin. The main concern of the commissions is to favor socially endangered households, as shown by the priority criteria. Finally and most importantly perhaps, any decision based on the origin of an applicant, that is, discriminating on this basis, is prohibited in France. Public housing offices are regularly audited: if evidence of discrimination is detected, they are judged and punished accordingly. This is why the lawyers Rouquette and Lipietz (1991) stress that the rules of allocation of public housing units that prohibit "localism," and the high administrative barriers that effectively prevent exchanges of lodgings except for changing spatial needs of families, make the allocation of public housing units largely exogenous with respect to the ethnic origins of the applicants.

<sup>15</sup> These associations are officially approved by the administrative head of the *département*, the *préfet*.

<sup>16</sup> Public housing allocation in Paris serves as a useful concrete example. We draw on the official audit of Observatoire du Logement et de l'Habitat de Paris (2011). Paris is a special case as it is, because of its size, a *département* as well as a city. The application form, the commission, and the allocation process thus take place in Paris at the city level. As of January 2010, there were 186,017 public housing dwellings in Paris. Public housing buildings are scattered across all Parisian areas, with a high concentration (69 percent) in six districts (the 13th, 14th, 15th, 18th, 19th, and 20th *arrondissements*). Within Paris, 48.7 percent of households are under the income ceiling and could be theoretically eligible. In practice, only households with very modest incomes apply (71 percent have an income lower than the minimum ceiling for all France, equivalent to €2,345 per month for a household with two children). On December 31, 2010, there were 121,937 ongoing applications, to be compared with 12,500 public housing units allocated over the year 2010. The breakdown of the households that were granted a public housing unit in 2010 is as follows: 67.7 percent came from precarious housing, 28.8 percent came from the private rental sector, and 2.3 percent came from the public housing sector. In the latter case, those are people who moved for a larger space following an increase in their household size (only 12 percent of the public housing dwellings have more than three rooms). The mobility rate (defined as the ratio of new entrants over the total number of public housing dwellings) is particularly low: it reaches 5.5 percent in 2010. It is formally possible to indicate a precise neighborhood in the application form, but in practice, very few applicants (6.6 percent) do provide this information. More than half of the 121,937 applicants (52.9 percent) did not mention any particular area at all, probably because of the fear of being rejected on this ground. Among those who indicated an area of preference, 91.2 percent mentioned the area where they were already living.

Despite this legal process of allocation, one might still be worried about the possibility of self-sorting of households that refuse the residential allocation proposed by the commission. In theory, households can refuse up to three offers. However, self-sorting, especially on ethnic characteristics, seems unlikely to be a common practice.<sup>17</sup> Residential mobility within the public housing sector is very low because of the strong shortage of vacant public housing dwellings. This makes it unlikely that the selected households could be really picky about the diversity of their neighborhood (see the study by Simon [2003]). Moreover, rents are considerably lower in public housing than in private housing, increasing the opportunity cost of moving and assuring low turnover: the mobility rate in the public housing sector is even lower than for recent owners. Using data from the 2002 HS, Debrand and Taffin (2005) give precise measures of the 2002 annual mobility rates: it amounted to 10.3 percent for new owners, 15.9 percent for tenants in the private housing sector, but only 9.9 percent for tenants in the public housing sector. While 9.9 percent may seem high, we show that when households move in a public housing block, they almost never achieve a placement in a less diverse setting (this is reported in the first section of the online appendix devoted to further tests of the identifying assumption). A corollary of subsidized rents and low mobility is low vacancy rates. Around 2.75 percent of public housing units were vacant (1.75 percent for vacancies of more than 3 months) in the early 2000s as reported by the 2002 HS and the Union Sociale pour l'Habitat (Social Union for Housing; <http://www.union-habitat.org/>) in its 2011 annual report ([http://www.union-habitat.org/sites/default/files/Données\\_statistiques\\_2011.pdf](http://www.union-habitat.org/sites/default/files/Données_statistiques_2011.pdf)). In comparison, the vacancy rate for all types of housing units, public and private alike, was 7.9 percent during this period.

This has obvious implications for waiting time. The Paris Region Public Housing Office (<http://www.drihl.ile-de-france.developpement-durable.gouv.fr>) provides information on the average waiting time in the *départements* of this region. In Paris, the average waiting time is 6 years for a one-room flat, 9 years for a two-room or three-room flat, and 10 years for a flat of more than four rooms. In the Hauts de Seine *département* (southwest of Paris), the average waiting time is 4 years, taking all types of flats together. The corresponding figure for the other *départements* of the Paris region (Seine Saint Denis, Val de Marne, Seine et Marne, Yvelines, Essonne, and Val d'Oise) is 3 years. The question and answer section of this website also indicates that “if you refuse an offer that is adapted to your situation, you will probably not get another offer before a long time period.” Importantly, for persons engaged in an emergency process in order to get

<sup>17</sup> In practice, the share of households refusing a public housing offer is not negligible, but we show that such behavior does not reflect selection based on preferences for or against diversity. See sec. 1.3 of the online appendix.

a public housing unit because of unfavorable living conditions, refusing a proposal excludes them from the emergency procedure. Corroborating anecdotal evidence beyond the Paris region comes from information gathered by the *Journal du Dimanche* (a generalist weekly French newspaper; <http://www.lejdd.fr/Societe/Social/La-grande-pagaille-du-logement-social-en-France-645151>) in December 2013. It requested from 15 cities the average waiting time to get a public housing unit there. The article reveals that “people already living in public housing often wait more than others,” as is explicitly indicated in the response forms from Lyon, Lille, and Nantes. For instance, in Nantes (the only city providing precise figures), public housing tenants wait, on average, 34.7 months versus 21.2 months for first-time applicants.

All of these considerations imply that households seeking public housing have very limited control over the precise place where they will be located or the degree of diversity within the block to which they will be assigned. This gives some initial support to our assumption that the distribution of households across public housing blocks is blind to households’ ethnic characteristics and to their preferences for diversity.

#### *B. Test of the Exogeneity of Diversity in Public Housing Blocks*

We now provide a more formal statistical test for our identifying assumption: we test whether, at the level of the block, diversity correlates with measures of the distribution of exogenous public housing building characteristics. We focus on building characteristics that are fixed in the sense that residents have no control over them. This strategy is very intuitive: if the assignments are random, then knowing the fixed characteristics of the block should reveal nothing about block-level diversity. More specifically, we run the following regression:

$$\text{DIV}_l = \alpha + \beta_1 Z_l + \text{FEmunic} + \varepsilon_l, \quad (2)$$

where  $Z_l$  are the exogenous characteristics of public housing in each block  $l$ . The correlation between these characteristics and block-level diversity ( $\text{DIV}_l$ ) is measured conditionally on municipality fixed effects ( $\text{FEmunic}$ ), which are the smallest geographic level after housing blocks that we can control for (France is divided into more than 36,000 municipalities). We use the 1999 census data to measure the following building characteristics, representative at the block level: the share of buildings with an elevator, the share of buildings with a sewage system, and the median number of apartments per building. One may, a priori, think that the median date of construction of the buildings, or any block characteristic related to a building’s age, could be considered as a fixed characteristic. The age of the building is actually highly correlated with the waves

of immigration and therefore with diversity, as we discuss at the end of this section. For this reason, note up-front that we do not perform the proposed test on variables related to a building's construction date, but that we use building age as a control variable in the main regressions of the paper. The regressions are run on the 2,492 blocks for which we observe public housing buildings in the HS.

The results of this test are presented in table 1. Columns 1–3 show the coefficients associated with each characteristic in separate regressions. All the characteristics are then included at once in column 4. The  $F$ -test for the null hypothesis  $\beta_1 = 0$  cannot be rejected at the 1 percent level for all coefficients of the three fixed characteristics that are expected to be totally unrelated to DIV. This simple test is in line with the idea that the public housing offices allocate dwellings to households without taking their origins into account, hence supporting our assumption of exogeneity of diversity in the public housing sector.

As mentioned above, any block characteristics related to building age could have been considered as a fixed characteristic over which residents do not have control if it was not highly correlated with the various waves of immigration. The oldest and first public housing units were built be-

TABLE 1  
REGRESSION OF DIVERSITY ON EXOGENOUS BUILDING CHARACTERISTICS

	DEPENDENT VARIABLE: NEIGHBORHOOD DIVERSITY BY NATIONALITY				
	(1)	(2)	(3)	(4)	(5)
Share of buildings with elevator	-.012 (.017)			-.027 (.017)	-.005 (.016)
Share of buildings with sewage		.064 (.071)		.014 (.095)	.049 (.101)
Median number of units			.001 (.000)	.001* (.001)	.000 (.001)
Median date of construction					-.019*** (.004)
Intercept	.232*** (.005)	.149** (.069)	.205*** (.005)	.214** (.092)	.258*** (.097)
Adjusted $R^2$	.444	.438	.442	.445	.465
Observations	2,155	2,492	2,450	2,137	2,137
$p$ -value	.487	.371	.118	.219	.731

NOTE.—Each column corresponds to a separate regression of neighborhood diversity on the block-level building characteristics listed in the left column. Each regression includes municipality fixed effects. Standard errors clustered at the municipality level are reported in parentheses. The last line reports  $p$ -values: in cols. 1–3, it corresponds to the significance test of each of the three fixed characteristics; in cols. 4 and 5, it corresponds to the  $F$ -test of joint significance of the the three fixed characteristics.

\*  $p < .10$ .

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

fore World War II to welcome both native French workers and immigrants from southern Europe (Italy and Spain in particular). The vast majority of public housing structures were built after World War II to welcome native French leaving the agricultural sector as they moved to industrial agglomerations. Shortly thereafter, these new structures housed the very large inflows of immigration from the Maghreb that took place between 1950 (postwar reconstruction period) and 1974 to help boost this industrial development. After 1974, the French government decided to reduce immigration drastically, limiting it to family reunification. Since mobility rates are almost nil in public housing, the ethnic composition is largely shaped by those various immigration waves. From the HS survey, we do observe a peak in the level of diversity in public housing buildings built between 1949 and 1974.<sup>18</sup> Therefore, when we regress block diversity on block characteristics related to the age of the buildings (median date of construction or share of buildings constructed after 1974), we unsurprisingly obtain a significantly negative coefficient. Yet, as explained, this does not come as a contradiction to our assumption. In addition, when we replicate the regression of column 4, controlling for the median date of construction, the  $p$ -value for the  $F$ -test of joint significance of the three fixed characteristics considered above is larger than .7, as reported in column 5. This suggests that diversity in the public housing sector is exogenous conditional on the age of the building, which is therefore a key variable to be included in our analysis.

Our identifying assumption is supported by a variety of alternative tests presented in section 1 of the online appendix, which we briefly summarize here. First, we run additional placebo tests but at the individual level. We estimate the effect of  $DIV_i$  on individual outcome variables that logically cannot be related to diversity, such as the perception of the quality of public goods that are financed and managed at a more aggregate level (e.g., by the municipality) rather than locally by the public housing offices. Likewise, we find that the  $DIV_i$  coefficient is not statistically significantly different from zero for those outcomes (sec. 1.1). Second, we test the exogeneity of the different steps in the allocation process during the application and the refusal decision process. We show the absence of any self-sorting along ethnic lines focusing on movers into public housing blocks (sec. 1.2). Since self-selection could still occur prior to the move, we also focus on households that have refused a public housing dwelling offer. We show that households having declined an offer end up living in public housing blocks that display the same level of diversity as those that directly accepted their first offer. Thus even if households

<sup>18</sup> The average level of diversity is 22 percent for buildings constructed before 1948 and 28 percent for buildings constructed between 1949 and 1974; it then continuously declined from .20 for buildings built between 1975 and 1981 to .15 for those built after 1999.

try to be choosy with respect to the ethnic composition of their neighborhoods, they eventually do not self-segregate in the public housing sector because of the allocation process and the tight supply constraints of dwellings (sec. 1.3). Finally, we compare the observed distribution of diversity across blocks within each *département* with a randomly simulated distribution of households and find that the equality between the two distributions cannot be rejected in most *départements*, supporting further our identification assumption (sec. 1.4). To sum up, all those tests point in the direction of diversity being exogenous in the public housing sector.

#### IV. Analysis

##### A. Specification

This section estimates the impact of diversity on social relationships and public goods at the local housing block level. We identify the causal impact of diversity by focusing on the public housing sector where households are exogenously allocated with respect to ethnic characteristics. Let  $j$ ,  $k$ , and  $l$  indicate, respectively, households, buildings, and blocks. For each outcome, we estimate the following equation:

$$Y_k = \alpha + \text{DIV}_l\beta + X_j\gamma + Z_k\delta + W_l\mu + \text{FEmunic} + \varepsilon_{jkl}, \quad (3)$$

where  $Y_k$  denotes the housing outcome we are interested in, as stated by household  $j$  living in building  $k$  and block  $l$ . Most of the outcomes we consider pertain to the building, but some of them refer to the neighborhood (in which case we consider a  $Y_l$ ). The variable  $\text{DIV}_l$  is the level of ethnic diversity in the block,  $X_j$  is a vector of household characteristics,  $Z_k$  is a vector of building characteristics, and  $W_l$  is a vector of socioeconomic characteristics of the block.<sup>19</sup>

The vector of controls includes first the household characteristics that are related to the selection criteria into the public housing sector ( $X_j$ ), in particular, household size and (log) household income per member.<sup>20</sup> Second, we control for block-level characteristics  $Z_k$  that could be confounding factors and be spuriously correlated with  $\text{DIV}_l$ . This includes the date of construction of the building (in six categories) since it might be a strong predictor of housing quality and explain part of the degradations observed. But this variable is also spuriously correlated with diver-

<sup>19</sup> Section 4 of the online appendix tests the robustness of our results to alternative specifications. First, we control for ethnic group shares in addition to fractionalization. Second, we try using an alternative measure of diversity based on a proxy for French speaking. Our main results remain unchanged in both cases.

<sup>20</sup> We have tested alternative specifications including age, gender, education, employment status, and nationality of the household head since those characteristics could also influence the opinion on housing conditions. Yet these variables have a very limited explanatory power and do not change our main results.

sity since it captures the different immigration cohorts, as suggested by the results presented in Section III.B. Third, we add two variables to control for the social and economic background of the neighborhood ( $W$ ). One is a detailed classification in 27 categories of the socioeconomic environment of each neighborhood, constructed by Tabard (2002) from the INSEE. This classification characterizes each neighborhood according to the socioeconomic category and the occupation of all male inhabitants. We use the classification that was built using the 1999 census data. This is the most detailed variable available in French national surveys to capture the socioeconomic background of a neighborhood. Indeed, an important issue is whether the degree of fractionalization is picking up various dimensions of the environment where people are living, including the extent of inequality and the unemployment rate or the socioeconomic background of the neighborhood (Alesina and La Ferrara 2002). The other is block unemployment rate (computed using the 1999 census) since this is a potential confounding factor for explaining criminality and other socioeconomic outcomes, as shown by Fougère, Kramarz, and Pouget (2009) and Hémet (2013). Finally, all the regressions include municipality fixed effects, which is the smallest geographic level after housing blocks that we can control for in French national databases. They correspond to *arrondissements* in large cities (Paris, Lyon, and Marseille) and are otherwise small cities. All results derive from ordinary least squares (OLS) estimates, with robust standard errors clustered at the housing block level.<sup>21</sup>

### B. *The Various Effects of Diversity*

The HS covers a large variety of questions documenting housing conditions, from social relationships with neighbors to the quality of the housing environment. Table 2 reports the descriptive statistics of our main outcomes of interest. To organize the discussion, we distinguish three main dimensions: (a) the neglect or voluntary degradations that are directly under the control of the tenants, that is, for which they can be held responsible; (b) the poor quality of basic housing facilities that are under the control of the public housing offices (the landlords) as a result of a lack of maintenance and repairs; and (c) public safety outcomes such as personal aggression and robberies, reflecting interpersonal or interethnic conflict. We have also run an exploratory factor analysis that yields similar, if not identical, categories. The results obtained with the three indices resulting from factor analysis are reported in section 1 of the online appendix.

<sup>21</sup> Logistic regressions on dummy outcomes yield similar results. To ease the interpretation of the coefficients, we will report the OLS estimates henceforth.

TABLE 2  
DESCRIPTIVE STATISTICS FOR EACH OUTCOME

	Mean	Standard Deviation	Range
A. Degradation of the Common Areas Due to Vandalism			
Damage to the premises	1.637	.778	1–3
Graffiti	.257	.437	0–1
Garbage on the floor	.188	.391	0–1
Broken windows	.136	.343	0–1
Broken doors	.127	.333	0–1
Broken lightbulbs	.094	.291	0–1
Broken mailboxes	.154	.361	0–1
Vandalism on the elevator	.085	.279	0–1
Noise in daytime	1.595	.748	1–3
Noise at night	1.374	.627	1–3
B. Poor Quality of Housing Due to Low Maintenance			
Low care of the common areas	1.593	.752	1–3
Poor condition of the façade	2.433	.962	1–5
Cold in the apartment	.175	.380	0–1
Cold due to bad insulation	.065	.246	0–1
Cold due to breakdown in heating equipment	.045	.207	0–1
Cold due to poor equipment	.059	.236	0–1
Poor quality of soundproofing	1.981	.823	1–3
Breakdown of the elevator	.155	.362	0–1
Toilet malfunction	.153	.360	0–1
C. Low Public Safety			
Robberies	.095	.293	0–1
Aggressions	.081	.273	0–1
Burglaries (or attempts)	.041	.198	0–1
D. Dissatisfaction with Housing Conditions			
Condition	2.502	.983	1–5

NOTE.—Depending on the outcome, we have between 4,310 and 5,189 observations.

The first dimension of housing quality refers to actions or goods that are largely under the control of the tenants. In this category, we include all the variables reporting neglect or voluntary deterioration in the common areas of the building. First, households are asked a general question on degradations: “Were the common areas of your building (lobby, staircase, floors) vandalized or neglected (destruction, deterioration) over the last 12 months?” The answers are 1 for “Never,” 2 for “Minor degradations,” and 3 for “Major or very frequent degradations.” Households are then asked to mention which kinds of degradations they observed over the previous year. They can choose several possible answers from the following list: graffiti or degradations of the walls (or on the floor), trash and

litter on the floor, broken windows, broken doors, broken lightbulbs, degradation of mailboxes, degradation of the entry phone or entry code, and deterioration of the elevator. For each outcome, the variable is coded as 1 in case of a degradation and 0 otherwise. All those items refer more or less directly to willful degradation. We will thus refer to this set of questions as the category “vandalism.” We also include in this category a question about noise pollution:<sup>22</sup> “How frequently are you disturbed by the noise in your housing during the day?” “During the night?” The answers are 1 for “Infrequently or never,” 2 for “Rather frequently,” and 3 for “Very frequently.”

The second main dimension of housing quality refers to goods that are not directly produced or altered by residents. But they might be related to diversity by the lack of maintenance and repairs by the public housing office to improve the housing quality. We will henceforth label this category “poor housing quality.” We include in it variables corresponding to housing problems that can be neither caused nor solved by the tenants but for which public housing offices are responsible. The households are first asked “How would you qualify the way the common areas of your building are maintained and taken care of (cleaning, maintenance of collective facilities: lighting, trash cans, . . .)?” The answers range from 1 for good to 2 for average and 3 for bad. More specific questions are also asked: “How does the façade of your building look?”<sup>23</sup> “What is the quality of the soundproofing of your housing?”<sup>24</sup> “Was the elevator out of order during more than 24 hours over the past three months?”<sup>25</sup> “Did you experience toilet issues (leaks, flush breakdown, drainage problems) over the last three months?” “Did you experience coldness in your apartment during more than 24 hours over the past 12 months?” We also include more detailed questions concerning the origin of coldness: “Did you experience coldness because of bad insulation?” “Did you experience coldness because the heating equipment broke down?” “Did you experience coldness because of poor heating equipment?” For all the previous questions, the variable is 1 when the answer is yes and 0 otherwise.

The last dimension of housing quality refers to interpersonal aggression and criminality. We will label this category “public safety.” Three questions correspond to this category: “Have you, or a member of your

<sup>22</sup> The underlying assumption is that the source of the noise in the hallways and apartments of the building is not due to poor soundproofing. As a matter of fact, we see in the following sections that more diversity increases the disturbance related to noise but fails to explain the quality of soundproofing.

<sup>23</sup> There are five possible answers: 1 = as new, 2 = good, 3 = average, 4 = dirty, 5 = bad, with cracks, and 6 = very bad, the building is threatened with collapse.

<sup>24</sup> The possible answers are 1 = good, 2 = average, and 3 = bad.

<sup>25</sup> In contrast, the question mentioned in the vandalism section refers to the interior status of the elevator rather than its mechanical breakdown.

household, been a victim of or a witness to physical aggression in your neighborhood during the last 12 months?" "Have you, or a member of your household, been a victim of or a witness to a robbery in your neighborhood during the last 12 months?" "Have you been a victim of a burglary (or any attempt) over the past 12 months?" For these three questions, the variable equals 1 in case of the event and 0 otherwise.

Finally, we look at the impact of diversity on the overall satisfaction about housing conditions, using the question "In general, how do you judge the quality of your housing conditions?" The variable takes on values from 1 for very good to 5 for very bad. This question on well-being related to housing conditions is rather general and summarizes in a way the various dimensions in which housing conditions could be affected by ethnic fractionalization.

### C. Results

Table 3 shows the effect of ethnic fractionalization on the various outcomes related to the three different dimensions—vandalism, poor housing quality, and public safety—and on the respondent's general level of satisfaction with housing conditions. For each outcome, we run a separate regression of the form of equation (3), controlling for household selection characteristics, age of the building, and the socioeconomic background of the neighborhood and including municipality fixed effects. The estimated coefficients for controls are presented in Appendix table B1.

Panel A of table 3 reports the effect of ethnic diversity on outcomes related to voluntary degradations and vandalism. For almost all the outcomes considered, the estimated effect of ethnic diversity is statistically significant at the 1 percent level and is sizable.<sup>26</sup> Take, for instance, the results for graffiti: a one standard deviation increase in ethnic diversity is associated with a rise by 9.2 percentage points in the probability of observing graffiti, which represents 21 percent of the total standard deviation of this outcome. The effect of ethnic diversity is twice as large as the effect of local unemployment: a one standard deviation increase in the block unemployment rate is associated with a rise by 5 percentage points in the probability of observing graffiti, which corresponds to 11.5 percent of a standard deviation of this outcome. Regarding deterioration, a one standard deviation increase in ethnic diversity induces a 4.2 percentage point increase in the probability of observing degradation of the elevator, which represents 14.9 percent of the total standard deviation of this out-

<sup>26</sup> The exception is for the indicator for broken doors and noise during the day, for which the effect of diversity is significant at only the 5 percent level.

TABLE 3  
THE EFFECT OF DIVERSITY ON HOUSING CONDITIONS: SUMMARY OF THE RESULTS

	ESTIMATED EFFECT OF NEIGHBORHOOD DIVERSITY ON ALL CONSIDERED OUTCOMES	
	Coefficient	Standard Error
A. Neglect in the Common Areas		
Damaged premises	.749***	(.169)
Graffiti	.510***	(.082)
Garbage on the floor	.391***	(.077)
Broken windows	.377***	(.070)
Broken doors	.185**	(.066)
Broken lightbulbs	.371***	(.059)
Broken mailboxes	.397***	(.076)
Vandalism on the elevator	.231***	(.062)
Noise in daytime	.316**	(.139)
Noise in nighttime	.452***	(.126)
B. Poor Quality of Housing		
Care of the common areas	.360**	(.155)
Façade condition	.192	(.194)
Cold in the apartment	.225**	(.076)
Cold due to bad insulation	.033	(.050)
Cold due to breakdown in heating equipment	.019	(.041)
Poor quality of soundproofing	.162	(.156)
Cold due to poor equipment	.098*	(.051)
Breakdown of the elevator	.264***	(.068)
Toilet malfunction	.058	(.066)
C. Public Safety		
Robberies	.049	(.049)
Aggressions	.003	(.050)
Burglaries	-.011	(.032)
D. Dissatisfaction with Housing Conditions		
Conditions	.335*	(.173)

NOTE.—The estimates presented in this table are obtained from the regression of each considered outcome using the new baseline specification. The controls are the following: date of construction of the building (six categories), neighborhood socioeconomic characteristics (unemployment rate and Tabard index), variables of selection in the public housing allocation process of the household (household size and annual income per equivalent household member, in log), and municipality fixed effects. The estimates for the other covariates are available on request. Standard errors clustered at the block level are reported in parentheses.

\*  $p < .10$ .

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

come. This effect is three times as large as that of the local unemployment rate.

The second set of regressions (panel B) shows the effect of diversity on outcomes signaling poor quality of housing. The coefficient associated

with ethnic diversity is positive and significant at the 1 percent level for problems related to elevator maintenance. It is also significant at the 5 percent level for the maintenance of the common areas and for coldness in the apartment. Diversity, however, has no significant impact on soundproofing quality and toilet malfunction. The estimated effects of diversity, while significant, are somewhat lower than those found for outcomes associated with vandalism. Consider the probability of the elevator being out of order. A one standard deviation increase in diversity is associated with a 4.7 percentage point increase in the probability that the elevator was out of order during at least 24 hours over the last 3 months, which is 13.1 percent of the standard deviation of this outcome. If we now turn to heating issues, our estimates imply that a one standard deviation increase in ethnic diversity is associated with a rise by 4 percentage points in the probability of experiencing insufficient heat in the apartment during more than 24 hours over the past year, which is 10.7 percent of the outcome's standard deviation. Panel C of table 3 reports the results for outcomes related to public safety, capturing direct aggression, robberies, and burglaries. Remarkably, it shows that ethnic diversity does not have a significant impact on any of these outcome variables. We offer an interpretation of this result in the next section.

We finally look at the general level of satisfaction with housing conditions. Panel D shows that ethnic diversity has a negative effect on satisfaction with housing condition, statistically significant at the 10 percent level. A one standard deviation increase in ethnic diversity generates an increase in the dissatisfaction with housing conditions that amounts to 6.1 percent of its standard deviation. To get a better sense of the magnitude of this effect, we can say that the increase in the dissatisfaction with housing conditions generated by a one standard deviation increase in the block unemployment rate corresponds to 15.5 percent of its standard deviation.

In sum, and taking advantage of data at a more micro level than has heretofore been available, we see that fractionalization operates with different degrees of impact for different sorts of public goods.<sup>27</sup> Fractionalization at the local level increases vandalism by a great deal, decreases building maintenance by a moderate (but overall significant) degree, and has no effect on security. The general negative impact of diversity on the various outcomes is thus consistent with the negative effect found on the general satisfaction with housing conditions. These findings allow us to propose in the next section the various channels through which fractionalization affects the provision of public goods.

<sup>27</sup> Our findings are unchanged with regressions on aggregated indices obtained with a principal component analysis and with a mean effect analysis, as shown in secs. 1 and 2 of the online appendix.

## V. The Channels of Impact of Ethnic Diversity at the Local Level

### A. *Rationalization of the Channels*

To rationalize our findings, we propose different interpretations of the channels through which fractionalization could affect local public goods related to living conditions and housing quality. The category of vandalism refers to the neglect or voluntary degradations of the common areas of the building, such as damage to common property, graffiti, or trash on the floor. These are outcomes over which public housing residents have control and for which they can be held responsible.<sup>28</sup> The category of quality of housing includes variables such as quality of soundproofing or coldness in the apartment. Those variables are more of the responsibility of the public housing managers. Finally, the public safety category represents outcomes that are less under the control of local public housing managers than of the police.

Our interpretation of the results on vandalism is that diversity prevents the creation of social norms to punish defectors, as the threat of social sanctions is lower across groups. Other-regarding preferences are less effective in more diverse areas. This has been a standard result in the literature since the seminal work of Coleman (1988), and it helps explain why we observe more voluntary degradations with diversity. In support of our intuition, many households living in the public housing sector do not generally report having “bad” or “very bad relationships” with their neighbors. Rather, they are more likely to report “no relationship at all” with their neighbors, which prevents the creation of other-regarding social norms. The increase in graffiti in more diverse areas might also illustrate the need to mark one’s territory in a context in which several groups coexist. In any case, cross-group sanctioning to prevent vandalism has been ineffective compared to what we see in more homogeneous blocks.

We understand the result on quality of housing as the inability of more heterogeneous communities to undertake collective action that would pressure the public housing office into improving housing quality. This could be sustained (though we have no direct evidence to support this) by beliefs in the housing directorate that it need not maintain public goods to high standards in heterogeneous housing projects because the likelihood of collective action against it is minimal. In this sense, the resulting poor housing quality associated with ethnic diversity can be seen as an equilibrium in which the lack of expectations of collective action would fail to give incentives to the housing directorate to make costly im-

<sup>28</sup> Given that residents need to enter a code in order to gain entry into their building, it is unlikely that these degradations are coming from outsiders.

provements.<sup>29</sup> The results concerning heating provide support for this assumption: we find not only that households living in more diverse neighborhoods report more heating failures but also that this is due to the poor quality of the heating equipment, an appliance typically under the control of the public housing office.

Finally, we can think of three possible interpretations of the absence of any diversity effect on aggressions and robberies. First, the data presented in the following subsection suggest that managers in more diverse environments invest more in security equipment than in building quality. Second, the absence of any impact of diversity on public safety outcomes could result from more physical security provided by a higher level of administration not subject to the constraints of local diversity, such as national or municipal police forces. To test this interpretation, we use data on municipal police, that is, police forces that are managed by the mayor at the municipality level, and hence are more closely related to local concerns.<sup>30</sup> Although the municipal police is only one part of the police force working in a given municipality (the rest of it being the national police), its main role is to maintain peace, security, safety, health, and tranquility in the municipality (art. L2212-5 of the general code of local territories) and is therefore more relevant for our analysis than national police. The Ministry of Interior has made available (on the government open-data webpage; <http://www.data.gouv.fr/fr/dataset/les-polices-municipales-par-commune>) the 2012 number of municipal police officers in the 4,202 municipalities providing this service. From this source, we can check how local police forces relate to the level of diversity at the municipality level. Table 4 presents correlations between the (log) number of municipal policemen per 1,000 inhabitants in a given municipality and the level of diversity in this municipality, measured in 2011. Column 1 shows that the raw correlation between the two is significantly positive: the more diverse a municipality is, the larger the number of policemen per inhabitant. The coefficient is still strongly positive in columns 2 and 3, where we progressively control for city size and *département* fixed effects. This evidence supports our second interpretation of a higher presence of police preventing criminal activities. A third explanation, one that encompasses findings for all three broad types of outcomes, would be that individuals living in the public housing sector experience social anomie. In fact, one-third (32.7 percent) of the public housing population, irrespective of diversity, declares to have no relationship at all with individuals living in their same neighborhood. In addition, we find that individ-

<sup>29</sup> The collective action could also influence the mayor's office. But the political logic of public housing support is beyond the scope of this paper.

<sup>30</sup> By contrast, the national police depend on the *préfecture*, at the *département* level. Information on national police at a very local level is not available.

TABLE 4  
MUNICIPAL POLICE OFFICERS AND DIVERSITY

	DEPENDENT VARIABLE: NUMBER OF MUNICIPAL POLICE OFFICERS PER 1,000 INHABITANTS (Log)		
	(1)	(2)	(3)
Municipality diversity by nationality	.559*** (.110)	1.368*** (.112)	.700*** (.145)
Number of inhabitants (log)		-.235*** (.012)	-.169*** (.012)
Intercept	-.882*** (.021)	1.027*** (.099)	.557*** (.099)
Adjusted $R^2$	.008	.113	.265
Observations	3,239	3,239	3,239
<i>Département</i> fixed effects	No	No	Yes

\*  $p < .10$ .

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

uals living in a more diverse neighborhood tend to have slightly fewer relationships with their neighbors.<sup>31</sup>

Given the nature of social relations in public housing, the most plausible interpretation is that diversity generates social anomie, that is, the absence of common rules and social norms. As a consequence of anomie, there are (a) weaker other-regarding preferences and a lack of credible threats of social sanctions, and hence more neglect and vandalism; (b) a failure to generate collective action to pressure the public housing offices into improving housing quality; and (c) fewer opportunities for violent confrontation where diversity might have increased incentives.

### B. Interpretation of the Channels Based on Repairs

We bring additional evidence about the channels based on the repairs performed in the building. The heterogeneous effect of diversity on repairs depending on the type of public good allows us to tease out the different channels through which diversity operates. Besides, maintenance and repair provide an objective interpretation to the previous subjective questions.

The HS asks whether elevators, staircases, windows, heating equipment, security equipment, and so on have been repaired or installed dur-

<sup>31</sup> Our findings are consistent with those of Fearon and Laitin (1996), who argue that despite interethnic relations being generally tenser, in-group policing mechanisms emerge endogenously that keep violence off of the equilibrium path. However, in this case, social anomie appears to be a more plausible mechanism sustaining peace as in-group policing is not likely to emerge where within-group collective action is not observed.

ing the previous year. We build three measures of repairs, corresponding to our three general outcomes. We define a first variable tracking repairs that usually come in response to degradation imputable to the tenants. Those repairs concern staircases, windows, doors, and lights in the common areas, that is, repairs related to neglect or voluntary degradations. A second variable indicates repairs that can be fixed only by the external intervention of the public housing office. Those repairs include revamping of the façade or interventions to improve, among other things, the heating system or insulation quality, that is, repairs related to the general quality of housing. Finally, we build a third variable accounting for the installation of security equipment in the building, which can be related to public safety outcomes. We then regress each of these three variables (as well as less aggregated indicators of repairs) on the level of diversity at the block level.

The OLS estimates are reported in panel A of table 5. Column 1 shows a positive and statistically significant correlation between the probability of repairs inside the building (windows, doors, and lights in the common areas) and local diversity. In the baseline regressions in Section IV.C (panel A, table 3), we already found that voluntary degradations increased with diversity. This implies that the larger number of repairs results from a greater need to fix things destroyed by the lack of publicly spirited social norms rather than from greater responsiveness by the housing authorities to regular maintenance.

Column 2 shows a negative and statistically significant correlation between the number of substantial works in the building (façade, heating, and insulation) and local diversity: the higher the diversity is, the less the work performed by the HLM offices for improving the general quality of housing. In the baseline regressions (panel B, table 3), we found that more diversity implies a lower quality of housing. This result thus suggests that more diverse neighborhoods are deprived of such substantial work, although the inhabitants actually complain (individually to survey enumerators) about the quality of housing. This supports our interpretation that tenants in more diverse neighborhoods are unable to engage in collective action to pressure the public housing offices into undertaking important works.

Finally, column 3 shows a positive and statistically significant relationship between the existence of security equipment and local diversity. In the main regressions (panel C, table 3), we found no impact of diversity on aggression and robberies. The presence of security equipment in more diverse neighborhoods can partly explain the absence of a diversity effect on burglaries, as mentioned in the previous subsection. Public housing managers may invest more of their budget in security equipment than in basic maintenance if they fear security threats.

TABLE 5  
RATIONALIZATION OF THE CHANNELS BASED ON REPAIRS

	REPAIRS RELATED TO		
	Vandalism (1)	Housing Quality (2)	Public Safety (3)
A. Relationship between Diversity and Type of Repairs			
Diversity	.131** (.054)	-.192** (.070)	.138*** (.040)
Adjusted $R^2$	.022	.007	.006
Observations	2,247	2,247	2,247
B. Correlation between Reported Degradation and Corresponding Repairs (%)			
Vandalism:			
Damaged premises	5.79***		
Graffiti	16.94***		
Garbage on the floor	14.35***		
Broken windows	11.86***		
Broken doors	13.74***		
Broken lightbulbs	12.24***		
Broken mailboxes	13.10***		
Vandalism on the elevator	13.54***		
Poor housing quality:			
Poor condition of the façade		-3.46***	
Cold in the apartment		3.89***	
Cold due to bad insulation		1.29	
Cold due to equipment breakdown		4.70***	
Cold due to poor equipment		.012	
Breakdown of the elevator		-.001	
Toilet malfunction		4.31***	
Low public safety:			
Robberies			2.52***
Aggressions			4.15***
Burglaries (or attempts)			2.35***

NOTE.—In panel A, the dependent variables displayed in the first line are defined as follows: (1) repairs related to vandalism: those likely to be done in response to neglect or voluntary degradation, i.e., repairs of staircases, windows, doors, and lights of the commons; (2) repairs related to poor housing quality: those taken care of by the public housing office, i.e., revamping of the façade, interventions to improve, among other things, the heating system, toilets, or insulation quality; (3) repairs related to low public safety: installation of security equipment in the building, i.e., entry code and locks. These three composite outcomes are regressed on block diversity, controlling for the number of dwellings in the building and its date of construction (in six categories). Panel B presents the correlation between each outcome variable and the corresponding composite measure of repairs, using the full HS sample. For instance, for broken windows, we report the correlation with repairs related to vandalism, while for robberies we report the correlation with repairs related to low public safety. Robust standard errors adjusted for block clustering are in parentheses.

\*  $p < .10$ .

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

## VI. Robustness Checks

The analysis has been based so far on self-reported subjective outcomes. A potential concern is that subjective perceptions could reflect personal bias rather than objective measures of public good provision. People might just be happier when they are surrounded by people more like themselves, and this is reflected in their answers to the quality of housing. We conduct several tests challenging this alternative explanation of personal bias.

First, we can exploit the various types of repairs and work done in the building or in the housing unit over the previous year. These variables have the advantage of being objective. Panel B of table 5 reports simple correlations between the various outcomes and the associated repairs. Note that because this test aims at testing for the relevance and objectivity of our outcome variables, it is performed on the full HS sample. We find that almost all the subjective outcomes are strongly and positively correlated with the existence of repairs, that is, objective outcomes, especially for the variables related to vandalism.<sup>32</sup> This is our first evidence of the reliability of our subjective measures of housing quality and well-being.

We then conduct more formal tests, reported in Appendix C. In the following we focus on the question about dissatisfaction with housing conditions, which is the most subjective of all the outcomes considered, but the other outcomes yield similar results. We start by regressing this variable on diversity with our baseline specification, and we add the respondent's nationality and interaction terms between diversity and nationality. We can test whether different groups react in different ways to the level of diversity of their neighborhood. Column 1 of table C1 shows that there is no different effect for the various groups, and the coefficient for diversity remains unchanged. Then we concentrate on differences between native French households' and fully Maghrebian households' dissatisfaction with housing conditions.<sup>33</sup> In particular, we interact the dummies of being in a fully native French household or being in a fully Maghrebian household with  $DIV_i$ : none of the coefficients is significant, as reported in column 2. Thus for any given level of diversity, there is no significant difference in the answers given by fully French and fully Maghrebian households. In other words, the idea that bad opinions of housing conditions are driven by average bad feelings due to being surrounded by foreigners can be rejected.

<sup>32</sup> An exception is the condition of the outside walls, which is negatively correlated with the probability that façade work was done. This is not surprising as the assessment of the façade's condition is provided at the time of the survey while repairs concern the previous year.

<sup>33</sup> Native French refers to both children and parents being born French in France.

We also look at within-housing project variation in perceptions by regressing self-reported perception on block fixed effects. Assuming there is no within-block variation in public goods, the remainder of the variation tells us if certain ethnic or sociodemographic groups are more likely to be positively or negatively biased. If perceptions have a high signal-to-noise ratio, there should be less within-block variation because perceptions would be a good signal of project-level public goods. Table C2 shows the results for dissatisfaction about housing conditions. Column 1 shows the within-block estimates by including block fixed effects. The only individual characteristic statistically significantly correlated with within-project variation in perceptions is household size. Income, education, age, and the country of origin of the household head are uncorrelated with perceptions of the environment. We also compute the standard deviation in the perception of the quality of housing between public housing projects and within-public housing projects. The standard deviation is almost twice as high across blocks (80.1 percent) as within blocks (43.5 percent), and this difference is statistically significant. This low level of within-block variation on perceptions adds confidence that there is an objective foundation for tenants' subjective reports.

We finally test the robustness of the previous results with respect to potential confounding factors in section 2 of the online appendix. We check how the inclusion of each control variable separately alters the effect of diversity on the various outcomes and test for an additional potential confounding factor: the average duration of tenancy of households in the block. On average, the various control variables do not significantly affect the estimated effect of diversity, suggesting little bias from confounding factors in our context.

## VII. Conclusion

This paper exploits French public housing policy as a natural experiment to identify the causal effect of diversity on the quality of local public goods related to well-being, living conditions, and housing quality. The French Housing Survey provides a unique micro level of analysis of social interactions among adjacent neighbors within housing blocks. We provide a detailed analysis of the channels through which diversity operates at the local level while the previous literature has focused so far on aggregate outcomes and channels. We use the exogenous allocation of households within public housing with respect to ethnic characteristics in France to address the bias from endogenous residential sorting that reduces the confidence in previous empirical findings on fractionalization. We find that fractionalization has a negative impact on other-regarding preferences, leading to higher neglect and vandalism in the housing commons. Fractionalization also undermines collective action

for the improvement of the quality of housing. But in our context, fractionalization has no effect on public safety, diversity being associated with social anomie within the housing blocks rather than violent confrontations among neighbors—helped as well by an increase in municipal policing in municipalities of high diversity.

This natural experiment calls for future research on the specific role of national, local, and informal institutions in mitigating or magnifying the effect of ethnic diversity on the provision of public goods. France is a country with a republican tradition that resolutely refuses to reify ethnic identification as a strategy to prevent the ethnification of everyday life. Yet we find a significant negative effect of diversity on local public goods in its public housing sector, comparable to the association found in the US localities where multiculturalist institutions regulate ethnic relations (Putnam 2007) and in cases in which public institutions are weak (Alesina and La Ferrara 2005). However, on issues of physical security in French public housing, the costs to ethnic diversity disappear. This may be due to the emergence of informal institutions (such as in-group policing as in Fearon and Laitin [1996]) or the supremacy of state-level institutions in which local diversity plays no role in the supply of order. Our evidence points to a possible third explanation, namely, that the social anomie resulting from diversity yields no contacts among neighbors rather than hostile ones. In any event, the results raise a puzzle, to be addressed in future research, on the general power of institutional arrangements in overcoming the negative implications of ethnic heterogeneity on the provision of public goods.

## Appendix A

### Descriptive Statistics

TABLE A1  
FRACTIONALIZATION BY NATIONALITY AT BIRTH  
FOR TENANTS IN PUBLIC HOUSING BLOCKS

	1999 Census	2002 Housing Survey
Mean	.28	.25
Median	.26	.23
Standard deviation	.19	.18
Minimum	0	0
Maximum	.875	.802

NOTE.—The average public housing tenant surveyed in the 1999 census lives in a block with 28 percent diversity. The average public housing tenant surveyed in the HS lives in a block with 25 percent diversity. The highest level of diversity observed in a public housing block is .875 in the census and .802 in the HS.

TABLE A2  
 SAMPLE CHARACTERISTICS: PUBLIC HOUSING POPULATION  
 (2002 Housing Survey)

	Mean	Standard Deviation
Household characteristics:		
Annual income per member <sup>a</sup>	12,226	7,923.83
Household size <sup>a</sup>	2.55	1.47
Household head characteristics:		
Age	47.09	17.13
Nationality		
Native French	.82	.38
Naturalized French	.06	.23
Maghrebien	.07	.26
Other African	.01	.12
Employment status:		
Employed	.58	.49
Unemployed	.11	.31
Inactive	.31	.46
Level of education:		
No diploma	.28	.45
Lower education	.51	.50
Baccalaureate	.09	.29
Higher education	.12	.32
Socioeconomic category:		
Craftsman, shopkeeper	.01	.12
Executive	.04	.19
Intermediate occupation	.12	.33
Employee	.20	.40
Blue-collar worker	.31	.46
Building characteristics:		
Number of units in the building	35.83	55.01
Date of construction of the building: <sup>a</sup>		
$t \leq 1948$	.06	.23
$1949 \leq t < 1974$	.52	.50
$1975 \leq t < 1981$	.15	.36
$1982 \leq t < 1989$	.11	.31
$1990 \leq t < 1998$	.12	.32
$1999 \leq t$	.04	.19
Block unemployment rate <sup>a</sup>	.22	.13

SOURCE.—Statistics were obtained from the 5,189 public housing observations in the 2002 Housing Survey.

<sup>a</sup> This variable is used as a control in our baseline specification.

## Appendix B

## Detailed Regression

TABLE B1  
 BASELINE SPECIFICATION: DETAILED REGRESSION

	Dissatisfaction with Housing Conditions
Diversity	.335* (.173)
Household selection variables:	
Income (log)	-.050 (.032)
Household size	.116*** (.012)
Date of construction of the building (ref: before 1948):	
1949 ≤ <i>t</i> < 1974	.039 (.092)
1975 ≤ <i>t</i> < 1981	-.086 (.098)
1982 ≤ <i>t</i> < 1989	-.105 (.098)
1990 ≤ <i>t</i> < 1998	-.443*** (.109)
1999 ≤ <i>t</i>	-.657** (.269)
Neighborhood characteristics:	
Block unemployment rate	1.175*** (.226)
Socioeconomic background (Tabard)	Yes
Intercept	2.279*** (.345)
Adjusted <i>R</i> <sup>2</sup>	.139
Observations	4,451

NOTE.—The coefficients correspond to the baseline specification, estimated for the dissatisfaction with housing conditions outcome. The regression controls for municipality fixed effects. Robust standard errors clustered at the block level are reported in parentheses.

\*  $p < .10$ .

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

## Appendix C

## Robustness Checks

TABLE C1  
ARE RESULTS DRIVEN BY SOME MAJOR ETHNIC GROUPS  
DISLIKING BEING AROUND FOREIGNERS?

	DEPENDENT VARIABLE: DISSATISFACTION WITH HOUSING CONDITIONS	
	(1)	(2)
Diversity	.414** (.187)	.369* (.209)
A. Nationalities (Reference: French at Birth)		
Naturalized French	.005 (.186)	
$DIV_i \times$ naturalized French	-.300 (.491)	
European	.048 (.232)	
$DIV_i \times$ European	-.036 (.651)	
Maghrebian	-.223 (.229)	
$DIV_i \times$ Maghrebian	.160 (.499)	
African	-.100 (.368)	
$DIV_i \times$ African	.531 (.886)	
Asian	-.102 (.989)	
$DIV_i \times$ Asian	-1.324 (1.650)	
Other nationality	1.163 (1.510)	
$DIV_i \times$ other nationality	-3.386 (6.227)	
B. Major Groups in HLM: Native French and Maghrebians		
Native French household		-.066 (.067)
$DIV_i \times$ Native French household		-.045 (.203)
Maghrebian household		.005 (.304)
$DIV_i \times$ Maghrebian household		-.200 (.668)
Intercept	2.347*** (.350)	2.296*** (.349)

TABLE C1 (Continued)

	DEPENDENT VARIABLE: DISSATISFACTION WITH HOUSING CONDITIONS	
	(1)	(2)
	B. Major Groups in HLM: Native French and Maghrebians	
Adjusted $R^2$	.140	.139
Observations	4,451	4,451

NOTE.—Each column presents the results of the regression of the dissatisfaction with housing conditions variables on different specifications. Each regression controls for the set of controls included in the baseline specification: household selection variables (income and size), date of construction of the building, block socioeconomic characteristics (unemployment rate and Tabard index), and municipality fixed effects. Robust standard errors adjusted for block clustering are in parentheses. For the sake of comparability, we remind that the diversity coefficient in the baseline specification is 0.335, significant at the 10 percent level.

\*  $p < .10$ .  
 \*\*  $p < .05$ .  
 \*\*\*  $p < .01$ .

TABLE C2  
 VARIATION IN PERCEPTION OF HOUSING QUALITY: WITHIN AND BETWEEN  
 PUBLIC HOUSING BLOCKS

	DEPENDENT VARIABLE: DISSATISFACTION WITH HOUSING CONDITIONS	
	Within Correlation (1)	Between Correlation (2)
Characteristics of the household head:		
Male	.017 (.065)	-.014 (.031)
Age	-.003 (.003)	-.003** (.001)
Diploma	.008 (.014)	-.031*** (.007)
Household income (log)	-.077 (.062)	-.146*** (.030)
Unemployed	.046 (.105)	.085* (.050)
Inactive	-.082 (.101)	-.063 (.047)
Household size	.092*** (.025)	.107*** (.012)
Naturalized French	-.111 (.114)	-.022 (.062)
European	-.028 (.177)	.105 (.086)
Maghrebien	-.089 (.135)	.019 (.061)
Other African	-.066 (.223)	.187 (.143)
Asian	-.342 (.790)	-.183 (.365)

TABLE C2 (Continued)

	DEPENDENT VARIABLE: DISSATISFACTION WITH HOUSING CONDITIONS	
	Within Correlation (1)	Between Correlation (2)
Other nationality	-.097 (1.197)	.409 (.520)
Building size (log)	.031 (.044)	.077*** (.010)
Intercept	3.002*** (.605)	3.618*** (.284)
Block fixed effects	Yes	No
Département fixed effects	No	Yes
Adjusted R <sup>2</sup>	.184	.081
Observations	5,188	5,188

NOTE.—Robust standard errors adjusted for block clustering are in parentheses.

\*  $p < .10$ .

\*\*  $p < .05$ .

\*\*\*  $p < .01$ .

## References

- Alesina, A., R. Baqir, and W. Easterly. 1999. "Public Goods and Ethnic Divisions." *Q.J.E.* 114 (November): 1243–84.
- Alesina, A., A. Devleeschauwer, W. Easterly, S. Kurlat, and R. Wacziarg. 2003. "Fractionalization." *J. Econ. Growth* 8 (June): 155–94.
- Alesina, A., and E. La Ferrara. 2000. "Participation in Heterogeneous Communities." *Q.J.E.* 115 (August): 847–904.
- . 2002. "Who Trusts Others?" *J. Public Econ.* 85 (August): 207–34.
- . 2005. "Ethnic Diversity and Economic Performance." *J. Econ. Literature* 43 (3): 762–800.
- Alesina, A., and E. Zhuravskaya. 2011. "Segregation and the Quality of Government in a Cross Section of Countries." *A.E.R.* 101 (August): 1872–1911.
- Bayer, P., S. L. Ross, and G. Topa. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *J.P.E.* 116 (December): 1150–96.
- Bernardot, M. 2008. *Loger les immigrés: La Sonacotra, 1956–2006*. Collection Terra. Bellecombe-en-Fauges: Éditions du Croquant.
- Caselli, F., and W. J. Coleman. 2013. "On the Theory of Ethnic Conflict." *J. European Econ. Assoc.* 11:161–92.
- Coleman, J. S. 1988. "Social Capital in the Creation of Human Capital." In "Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure," supplement, *American J. Sociology* 94:S95–S120.
- Combes, P.-P., B. Decreuse, B. Schmutz, and A. Trannoy. 2012. "The Neighbor Is King: Customer Discrimination in the Housing Market." Working Paper no. 1224 (September), Aix-Marseille School Econ.
- Dahlberg, M., K. Edmark, and H. Lundqvist. 2012. "Ethnic Diversity and Preferences for Redistribution." *J.P.E.* 120 (1): 41–76.
- Dancygier, R. M. 2010. *Immigration and Conflict in Europe*. Cambridge: Cambridge Univ. Press.

- Debrand, T., and C. Taffin. 2005. "Les facteurs structurels et conjoncturels de la mobilité résidentielle depuis 20 ans." *Économie et Statistique* (381): 125–46.
- Dunning, T., and L. Harrison. 2010. "Cross-Cutting Cleavages and Ethnic Voting: An Experimental Study of Cousinage in Mali." *American Polit. Sci. Rev.* 104 (1): 21–39.
- Easterly, W., and R. Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Q.J.E.* 112 (November): 1203–50.
- Esteban, J., and D. Ray. 2011. "A Model of Ethnic Conflict." *J. European Econ. Assoc.* 9 (3): 496–521.
- Fearon, J. D., and D. D. Laitin. 1996. "Explaining Interethnic Cooperation." *American Polit. Sci. Rev.* 90 (4): 715–35.
- . 2000. "Violence and the Social Construction of Ethnic Identity." *Internat. Org.* 54 (September): 845–77.
- Fougère, D., F. Kramarz, and J. Pouget. 2009. "Youth Unemployment and Crime in France." *J. European Econ. Assoc.* 7 (5): 909–38.
- Glennester, R., E. Miguel, and A. D. Rothenberg. 2013. "Collective Action in Diverse Sierra Leone Communities." *Econ. J.* 123 (568): 285–316.
- Goux, D., and E. Maurin. 2007. "Close Neighbours Matter: Neighbourhood Effects on Early Performance at School." *Econ. J.* 117 (523): 1193–1215.
- Hémet, C. 2013. "The Local Determinants of Victimization." Working Paper no. 1349 (October), Aix-Marseille School Econ.
- Jacquot, A. 2007. "L'occupation du parc HLM: Éclairage à partir des enquêtes logement de l'INSEE." Document de Travail no. F0708, Direction des Statistiques Démographiques et Sociales, Paris.
- Jha, S. 2013. "Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia." *American Polit. Sci. Rev.* 107:806–32.
- Katz, L. F., J. R. Kling, and J. B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Q.J.E.* 116 (May): 607–54.
- Kling, J. R., J. B. Liebman, and L. F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.
- Luttmer, E. F. P. 2001. "Group Loyalty and the Taste for Redistribution." *J.P.E.* 109 (June): 500–528.
- Miguel, E. 2004. "Tribe or Nation? Nation Building and Public Goods in Kenya versus Tanzania." *World Politics* 56:328–62.
- Miguel, E., and M. K. Gugerty. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya." *J. Public Econ.* 89 (December): 2325–68.
- Observatoire du Logement et de l'Habitat de Paris. 2011. "L'accès au logement social à Paris: Analyse de la demande de logement social et bilan des propositions et des attributions aux logements sociaux à Paris en 2010." Observatoire, Paris.
- Oreopoulos, P. 2003. "The Long-Run Consequences of Living in a Poor Neighborhood." *Q.J.E.* 118 (November): 1533–75.
- Page, S. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton, NJ: Princeton Univ. Press.
- Posner, D. N. 2004. "The Political Salience of Cultural Difference: Why Chewas and Tumbukas Are Allies in Zambia and Adversaries in Malawi." *American Polit. Sci. Rev.* 98 (11): 529–45.
- Putnam, R. D. 2007. "E Pluribus Unum: Diversity and Community in the Twenty-First Century; the 2006 Johan Skytte Prize Lecture." *Scandinavian Polit. Studies* 30 (2): 137–74.

- Rouquette, R., and H. Lipietz. 1991. *Droit du logement social organismes d'HLM, SEM, ANAH, PACT, ARIM*. Collection L'actualité juridique. Paris: Éditions du Moniteur.
- Simon, P. 2003. "Le logement social en France et la gestion des 'populations à risques.'" *Hommes et Migrations* (1246): 76–91.
- Tabard, N. 2002. "Inégalités et disparités entre les quartiers en 1999." In *Données sociales: La société française*. Paris: INSEE.
- Varshney, A. 2003. *Ethnic Conflict and Civic Life: Hindus and Muslims in India*. Oxford: Oxford Univ. Press.

# Online Appendix to:

## The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation

YANN ALGAN\*, CAMILLE HÉMET†, DAVID D. LAITIN‡

February 9, 2015

In this on-line appendix, we propose supplementary material that constitutes an important part of the research leading to the main paper. We provide additional statistical evidence supporting our identifying assumption in Section 1 and we test the robustness of our results in Section 2.

### 1 Additional tests on the exogeneity of residential allocation in the public housing sector

This section provides additional tests on the exogeneity of residential allocation within the housing sector. First we carry out placebo tests by estimating the impact of  $DIV_l$  on outcome variables that should not be in principle related to diversity (section 1.1). Second, we test the exogeneity of the different steps of the allocation process during the application and the refusal decision process (sections 1.2 and 1.3). A final test shows that the observed spatial distribution of diversity across housing blocks within *départements* is not significantly different from a random distribution obtained with Monte Carlo simulations (section 1.4). These tests all support results provided in the published article demonstrating the exogeneity of spatial allocation across housing blocks of a given locality in the French public housing sector.

We use the French Labor Force survey (*Enquête Emploi*, INSEE, hereafter LFS) to test our identification assumption that spatial allocation in the public housing market can be considered as quasi-random and exogenous relative to ethnic characteristics. The survey covers the period 2003-2007 and provides all the relevant information about ethnic background, economic characteristics and geographic location of individuals to test our identification strategy. The sampling unit of this database is at the housing block level, which consists of twenty adjacent households on average. Over the 2003 to 2007 period, more than 10,000 different housing blocks were sampled. All the households within a randomly selected housing block were surveyed and, within each household, all persons aged fifteen or over were interviewed.<sup>1</sup> Using these data, we can analyze real neighborhoods

---

\*Department of Economics, Sciences Po

†Institute of Economics of Barcelona

‡Department of Political Science, Stanford University

<sup>1</sup>INSEE has chosen this sampling strategy so as to reduce the traveling expenses of those who administer the survey.

at a very small geographic level. We also have information on whether the respondent was living in a public housing unit, whether he or she has been living in his or her current public housing unit for at least one year or whether he or she has just moved into the neighborhood. These particular features enable us to compute the level of ethnic diversity prevailing within each public housing block and to test for the absence of self-sorting on ethnic background across public housing blocks.

## 1.1 Placebo tests

We start by running placebo tests to show that housing units are allocated randomly within districts. We run the baseline specification on outcome variables that logically cannot be related to  $DIV_i$  and test for the null hypothesis that the associated coefficients are not statistically different from zero. The placebo tests are similar in spirit to the one reported in section 3.1 of the article, but we run this analysis at the individual level using questions from the HS.

The housing survey provides two sets of variables that we can use as placebos to perform the proposed test: fixed housing characteristics that should not be affected by diversity, and neighborhood amenities that are not under the control of HLM offices. The fixed housing characteristics include the following variables: whether the building has a collective gas supply as opposed to individual canisters; whether the ceiling of the main room is lower than three meters; whether most of the windows have double glazing; the extent to which the electrical installation is recessed (1: all the installation is recessed, 2: the entire installation is not recessed, but a part of the installation is protected by rods or tubes, 3: the entire installation is not recessed, and some wires are not even protected by rods or tubes), and whether there is electrical grounding. We regress these variables on diversity, using the full specification. The estimated effects of diversity on this first set of placebo variables are reported in column 1 of Table 1: none of those fixed housing characteristics are statistically significantly correlated with diversity, in line with the results presented in Table 1 (section 3.1) of the article.<sup>2</sup>

The second set of placebo variables relates to neighborhood amenities that are generally provided by a higher level of administration (e.g. municipality or département) rather than locally financed by the public housing offices. Since those goods or services are fixed and out of reach of the tenants or landlords, we should expect no effect of diversity. Besides, this is also a good test for potential reporting bias in general, since the effect on all types of public goods should be the same in this case. If it is related to localized collective action failures, then the impact should only be on locally provided/maintained public goods. The HS documents four public goods that enter this category: quality and maintenance of roads and streets, access to public transportation, accessibility of the area by private transportation, and quality of green spaces. For each variable, the values range from 1 for good, 2 for neither good nor bad to 3 for bad. The estimated effects of diversity on this second set of placebo variables are reported in column 2 of Table 1. We find that ethnic fractionalization is not correlated with these neighborhood amenities, except for the accessibility of the area by private transportation, though in this case only significant at the 10% level. These placebo tests confirm that diversity is not related to any fixed building characteristics or to public goods that are not

---

<sup>2</sup>We do not consider the age of the building as a placebo variable here given that it is correlated with the timing of the different immigration waves, and thus with diversity, as already shown and explained in section 3.1.

under the control of the public housing managers.

## 1.2 Absence of self-sorting

We now provide alternative tests on the exogeneity of the allocation process. Our first test consists in showing that households are unable to self-segregate in the public-housing market. We test this using the LFS subsample of all individuals who moved into an area within the previous year (new movers), whether in the public or private markets: this enables us to compare their individual characteristics with the diversity and ethnic composition of the neighborhood prior to their moving.

We first use this methodology to estimate the link between the new movers' nationality and the share of the area's "long term" population of the same origin.<sup>3</sup> Table 2 reports the results from an OLS regression of the share of neighbors from the same origin as new movers on new movers' characteristics: nationality group, public housing dummy, quadratic function of age, (log) hourly wage, education, socio-economic category, *département* fixed effects, and interaction of individual characteristics with the public housing dummy. We consider seven different nationality groups: native French, naturalized French, Europeans, Maghrebians, other Africans, Asians, and other nationalities, which is taken as the reference group. First, we find that the coefficient for living in the public housing sector is negative and significant at the 5% level: HLM households move in areas where the share of individuals from the same origin is on average 15.5% lower than for households in the private housing sector. Second and more importantly, none of the interacted terms between nationalities and public housing are statistically significant (except for naturalized French citizens but they don't constitute an ethnic group, and the effect washes out when we control for the share of one's own ethnic group in the *département*).<sup>4</sup> This comforts us with the idea that there is no self-segregation along ethnic lines in the public housing sector.

We conduct the same kind of test on other individual characteristics such as education, and reach similar conclusions. We do not find any statistically significant self-segregation along education levels in public housing blocks.<sup>5</sup> We also regress the probability of having moved in a new HLM dwelling in the past year (dummy equal to one in this case and to zero if the individual was already living in the same HLM apartment one year before) on individual characteristics (nationality, age, gender, wage, education, socio-professional group), and the interaction of these characteristics with the ethnic diversity among the public housing population of the block. All the coefficients on the interaction terms are not significantly different from zero, rejecting self-selection along all those characteristics. In the same vein, we estimate the correlation between new movers' hourly wages

---

<sup>3</sup>A similar test was proposed by Goux and Maurin (2007) to show that the educational achievement of the children of newcomers in public housing is uncorrelated with that of the current residents. Individuals do not self-select in public housing neighborhoods according to the educational achievement of the neighbors' children. By contrast, the authors find a strong self-selection on the educational characteristics in the private housing sector.

<sup>4</sup>Since the results might also be driven by the French population which constitutes the largest group, we have alternatively run the same regression on the subsample of foreigners only. None of the coefficients of interest (individuals' nationality, and its interaction with the public housing dummy) are significant. Yet, these results are based on a sample of slightly more than 600 individuals, and are insufficient to report as a main finding.

<sup>5</sup>The only characteristic for which we find a positive correlation between new movers and long-run inhabitants is the fact of being a factory worker. This should not come as a surprise since this occupation represents more than 30 percent of the public housing population: historically, public housing was initially (and over several decades) dedicated to factory workers.

and the level of diversity of the area into which they just moved. Without prior beliefs over agents' preferences, if individuals have a taste for or against homogeneity, there should be a significant relationship between the level of ethnic diversity prevailing in their neighborhood and their wages if their choice were not constrained by legal rules. We run OLS estimates of the hourly wage of newly arrived individuals into a public housing block on the level of diversity of the area in which they just moved, controlling for the municipality of residence. We do not find any significant correlation between the income of individuals moving into a public housing dwelling and pre-existing diversity of nationalities within the neighborhood.<sup>6</sup> Finally, we also exploit the information about the intensity and quality of relationships with the neighbors, provided by in the Housing Survey.<sup>7</sup> We find that public housing neighborhoods are characterized by an absence of any relationship, and by a lower quality of relationships when they exist. This brings additional support to our assumption that public housing tenants do not choose their neighbors; otherwise we would expect them to report more frequent relationships, of better quality.

### 1.3 Tests on the refusal rate of public housing offers

The previous tests point out the absence of self-selection across public housing blocks among the movers. But self-selection could occur prior to the move. We address this issue by looking at households that have refused an offer for a public housing dwelling. Actually, a non-negligible share of households waiting to be allocated into a public housing unit report to have declined at least one offer.<sup>8</sup> In principle, households can decline up to three offers. But due to the frustrating shortage of public housing dwellings, we show in this section that households that declined an offer in the past cannot self-select in the future into neighborhoods with different levels of diversity than those that did not refuse any offer.

First, if there were self-selection upon diversity, we should expect households that turned down proposals before being allocated to their current public housing dwelling to end up living in neighborhoods with different levels of diversity than those who did not. The bias could go in either direction, depending on the mechanism driving the relocation of these households. Suppose for instance that there are high- and low-diversity neighborhoods, that offers are random, and that half

---

<sup>6</sup>We find a very strong negative relationship between income and diversity for individuals having just moved into a private dwelling (the estimated coefficient is -0.14 and is significant at the 1 percent level). This is powerful evidence of the bias introduced with endogenous sorting in the previous literature on ethnic diversity.

<sup>7</sup>We use the following questions: "*How would you qualify the relationships with your close neighbors?*" and "*How would you qualify the relationships in your neighborhood?*", where the first question refers to the direct neighborhood, while the second one refers to a larger neighborhood. The possible answers are 1 for "Good", 2 for "Average", 3 for "Bad" and 4 for "No relationship". From this, we build two dummy variables indicating whether the individual reports any relationship (good, average or bad) in the direct or in the larger neighborhood, respectively. For individuals reporting having relationships, we built two additional variables characterizing the quality of the relationship in the direct or in the larger neighborhood, respectively.

<sup>8</sup>In the Housing Survey, 24.2 percent of households currently living in a public housing dwelling report to have rejected at least one proposal before finding their current place; 47.9 percent of these households declared that one of the major reasons for this decision was they found the local environment unpleasant. (The other possible answers were: inconvenient place, rent too expensive, low quality building, and apartment not corresponding to household needs.) The answer "unpleasant local environment" is hard to interpret though, since it could refer to diversity as well as the proximity to public transport and infrastructure, lack of green spaces and so on. Also hard to interpret are the 16.5 percent of the households that are still waiting for an offer at the time of the survey - whether they are already living in a public housing dwelling or not- who have previously turned down at least one offer. See below for a test on whether "unpleasant environment" was driven by a distaste for diversity.

of offers are for low-diversity neighborhoods. If people can refuse at most one offer, they will refuse the first one if and only if it is a high-diversity neighborhood. In this simple model, all the people accepting the first offer will live in a low-diversity neighborhood, and exactly half of people who refused the first offer will live in a low-diversity neighborhood. In other words, households taking the first offer are positively selected, while those that turned down the initial offer (and accepted the final one) are actually allocated randomly, so that they will live in more diverse neighborhoods. If, on the other hand, the market is a bit more flexible than in the previous example, and if households can refuse offers up to the point they are allocated to a low diversity block, then they should end up living in less diverse neighborhoods than those who immediately accepted.

To test the conjecture that refusals might be driven by selection based on diversity, we run OLS regressions of a variable indicating whether the household declined at least one offer (during the latest application process) on the level of diversity of the neighborhood in which it now lives.<sup>9</sup> Panel A-I of Table 3 shows various estimates of the effect of ethnic diversity on the probability of having turned down offers across our different specifications of control variables. In each specification, the coefficient on  $DIV_i$  is not significantly different from zero, showing that households having declined offers during their past allocation process do not end up living in neighborhoods with significantly different levels of diversity.

We then explore further the validity of this conjecture by focusing on the reasons adumbrated by households for refusing an offer. If public housing residents were to sort themselves on the basis of their (dis)taste for diversity, those who declined "because of the local environment" could now live in neighborhoods with a significantly different level of diversity. We regress a dummy variable indicating whether a household declined an offer due to an "unpleasant environment" (during the past application process) on the level of diversity of its current neighborhood. Panel A-II of Table 3 shows that none of the  $DIV_i$  coefficients is significant. Instead, household characteristics such as the labor market status of the head of household and the size of the household are the only ones that matter in these regressions.

Alternatively, we perform these tests on the subsample of individuals currently waiting for an HLM offer. Panel B of Table 3 shows the regressions of the refusal dummy (B-I) and the "refusal due to unpleasant environment" dummy (B-II), for individuals who are currently applying for public housing on the diversity in their neighborhood, controlling for the same set of characteristics. Once again, we find that the ethnic diversity of a block is uncorrelated with households wait-listed for an HLM assignment having turned down offers since the beginning of their request (B-I). This suggests that the current level of diversity in the block does not rush households out of the area, as their propensity to decline an offer is independent of  $DIV_i$  in the current neighborhood. The high refusal rates of HLM offers do not therefore seem driven by a hope to reduce diversity by waiting.<sup>10</sup>

---

<sup>9</sup>Probit estimates yield similar results.

<sup>10</sup>We ran the same test on the subsample of individuals who left their previous housing unit because they did not like the environment: while households that left their previous housing to escape from an unpleasant environment now live in blocks where the diversity is significantly lower in the private housing sector, this result does not hold for households that moved within the public housing market. This suggests once again that HLM households do not have control over the diversity of the block to which they are allocated.

## 1.4 A Monte-Carlo simulation-based test of exogeneity

This section presents a final test on the exogeneity of residential allocation in the public housing sector, which consists in comparing the observed distribution of diversity by nationality across public housing blocks with a random distribution. This is an alternative test of our assumption of quasi-randomness of the allocation of households across public housing blocks within a given area. We use Monte Carlo simulation to reproduce an artificially random distribution of the population. We randomly reallocate the public housing population across the different blocks within each *département*, and then compare this random distribution to the actual distribution.<sup>11</sup>

We use the Labor Force Survey to perform this test, because its unique design allow us to work on the entire population of the surveyed neighborhoods. For each *département*, we pool the public housing population and reallocate it randomly, without replacement, across the different residential blocks of the corresponding *département*, maintaining unchanged the actual size of each block. We get a simulated random allocation of individuals from different ethnic origins across blocks. We then run a Kolmogorov-Smirnov (KS) test of equality of distribution of ethnic fractionalization across neighborhoods with its actual distribution. We finally determine the percentage of *départements* for which the actual and simulated distributions across housing blocks are similar, i.e. those for which we cannot reject the null hypothesis of equality of the distribution at the 10% level. The Monte-Carlo simulation results presented here are based on one hundred replications of the process described above. For each draw we run the tests for the equality of distributions, and then average the results.

The equality of spatial distribution between the randomly simulated distribution of diversity and the observed one cannot be rejected for 73 percent of draws in the public housing sector. This is lower than the expected threshold of 90 percent of the draws with a test at ten percent, and we thus reject random allocation more often than one would expect if the true allocation were random. These results indicate that the allocation may not be completely random according to this test. Yet, the allocation process across public housing blocks is, by far, more random than the allocation in the private sector. When we run the same test within the private rental sector, the fraction of draws where equality cannot be rejected drops dramatically to 34 percent.

Figure 1 illustrates this difference by showing the observed and average simulated distributions of diversity in the private rental sector and in the public housing sector. The observed distribution of diversity (left-hand side graphs) is quite concentrated in the low diversity zone for the private

---

<sup>11</sup>As mentioned in section 2.1 of the paper describing the public housing allocation process, the allocation of households across public housing blocks takes place at the *département* level. If the members of the commission follow the legal criteria and do not take into account the ethnic characteristics in the allocation process, we should find an equal distribution of households of a given nationality across the various public housing blocks within each *département*. For the sake of illustration, let us assume that 10 percent of Maghrebians live in the public housing sector in Paris. We should find the same share of 10 percent of Maghrebians within each Parisian housing block if the allocation was truly exogenous with respect to ethnic characteristics. Naturally, this equality of distribution of ethnic group shares across housing blocks can hold only if we have a sufficiently large number of individuals within each housing block. Instead, in the Labor Force Survey, on which we perform the test, we only observe an average of 40.55 different individuals corresponding to 15.53 households in each neighborhood. This is due to the sampling strategy of the French National Institute of Statistics and Economics (INSEE) that interviews all the individuals from a given neighborhood, but consequently limits the size of the neighborhood. With such a small sample size of observations at the neighborhood level, any analytical test of equality of distribution of ethnic group shares across blocks would most likely fail.

housing blocks (upper-left corner), while it is almost flat in the public housing blocks (lower-left corner). This is consistent with our identifying assumption. It is also striking that a non-negligible number of blocks in both housing sectors have a zero diversity. This is to a large extent explained by the composition of the population (86 percent French) and the quite small size of neighborhoods in the LFS.<sup>12</sup> The right-hand side graphs plot the corresponding simulated distributions. As revealed by the KS test, the average random distribution of diversity across public housing blocks (lower-right corner) does not differ much from the observed one. By contrast, the simulated distribution in the private sector is much flatter in the 0 to 20 percent range of diversity than the observed one, as would be expected from the tests results.

## 2 Robustness tests

### 2.1 Sensitivity to control variables

This section tests the robustness of the previous results with respect to potential confounding factors. We check how the inclusion of each control variable separately alters the effect of diversity on the various outcomes. To do so, we regress the outcomes on  $DIV_i$  following the main equation of the paper, first without any control, as a reference, and then including each control variable alternatively. Table 4 reports the corresponding estimates of diversity, without any control in column 1, and with each control separately in columns 2 to 6. In column 7, we test for an additional potential confounding factor: the average duration of tenancy of households in the block. Indeed, if the ethnic composition of the pool of applicants for public housing changes over time, diversity could proxy for the average time that tenants have lived in the block. Blocks with longer-term residents may thus have more social capital, and hence different outcomes.<sup>13</sup>

This table also reports the 95 percent confidence intervals for each estimated coefficient of  $DIV_i$  to check whether it is statistically significantly affected by the controls. On average, the various control variables do not significantly affect the estimated effect of diversity, suggesting that our results are not sensitive to the control variables. The only control variable that affects the magnitude of the diversity coefficient for some outcomes is the block unemployment rate. This is notably the case for "noise during daytime", which is consistent with the fact that blocks where the unemployment rate is higher are those where more people are likely to stay home longer, with an increased probability of being noisy or of being disturbed by surrounding noise pollution (works in the streets, trucks, and cars). The main impact of the unemployment rate is on aggression, with the  $DIV_i$  estimate losing its statistical power. This result confirms previous results on the effect of unemployment on criminality (Fougère et al., 2009; Hémet, 2013). By contrast, unemployment rate has no effect on the quality of housing, which provides additional support to the robustness of our results. Otherwise, we do not find any significant change in the  $DIV_i$  indicator across specifications, which suggests little bias from confounding factors in our context. These results provide additional

<sup>12</sup>We keep the zero diversity blocks in our main analysis; deleting them does not affect the results.

<sup>13</sup>Although this is not reported in the table, note that the effect of tenancy is significantly negative for most of the considered outcomes, and especially those related to vandalism, which confirms the idea of average tenancy as a proxy for social capital.

evidence that  $DIV_i$  is randomly assigned since the coefficient on  $DIV_i$  should not be sensitive to the inclusion of control variables in this case.

## 2.2 Fractionalization and ethnic shares

In this section, we test the robustness of our results to alternative specifications, for instance including the shares of each origin group, and also to alternative measures of diversity.

The basic regressions measure ethnic diversity using a standard fractionalization index, controlling for household, building and neighborhood characteristics. Yet, as suggested by Vigdor (2002), it might be important to control for ethnic group shares to get a more comprehensive set of covariates for diversity. Column 1 of Table 5 reports the results once we control for ethnic group shares.<sup>14</sup> The estimated impact of  $DIV_i$  is now even stronger than in the previous specifications, confirming the robustness of our result along this dimension.

Moreover, we run regressions replacing the fractionalization index by ethnic group shares (Column 2 of Table 5), and by ethnic group shares and their square (Column 3 of Table 5), controlling for the usual individual and local characteristics. Only one group (Maghrebians) seems to have a significantly negative effect on the dissatisfaction with housing conditions: the higher the share of Maghrebians in a block (relative to the share of French), the more likely individuals are to complain about their housing conditions. However, the marginal effect decreases with the share of Maghrebians. From this result, we infer that our measure of diversity reflects not only the actual ethnic composition of the neighborhood, but also that some ethnic groups might have different effects on self-reported perceptions of the quality of public spaces as they become a majority of the neighborhood population. However, this result does not call into question the effect of diversity per se on which we have already reported (Column 1 of Table 5).

Finally, we re-run our main regressions using an alternative fractionalization index, trying to encompass another dimension of diversity. More precisely, in order to account for communication issues potentially related to the diversity of origins, we compute an alternative fractionalization index based on a proxy for French speaking. We use information provided by the *Organisation Internationale de La Francophonie* to group countries according to the share of the population that is French speaking (in 2010). Their classification allows us to distinguish among six groups of countries: countries which are not members of the organization,<sup>15</sup> countries in which French speakers represent less than 5 percent of the population, countries in which French speakers represent between 5 and 15 percent of the population, countries in which French speakers represent between 16 and 35 percent of the population, countries in which French speakers represent between 36 and 60 percent of the population, and countries in which French speakers represent more than 60 percent of the population. We then assign one of the six French-speaking levels to each individual (from the 1999 census dataset), according to his or her nationality at birth. This sorting of individuals captures the probability that they actually speak French. Finally, we compute a standard fractionalization

---

<sup>14</sup>In Table 5, we aggregate the various nationalities at birth into six different categories. The results are similar when we work with more detailed shares for all nationalities. The share of native French is the omitted category.

<sup>15</sup>For these countries, the organization does not provide any data, but we can reasonably assume that the share of French speaking population in non-member countries is close to zero.

index for each block (using the same methodology as for our main index of diversity) relying on the shares of the block population belonging to one of the six groups. We replicate the regressions of the paper<sup>16</sup> using this alternative index instead of the one based on nationality at birth. The results are reported in Table 6. Each coefficient comes from a separate regression. The columns indicate the four dependent variables under study. Panel A and panel B respectively correspond to the measure of diversity used in each regression. Our results are unaffected when we use the new index based on French speaking origin. This is not very surprising given that the correlation between the two indices is very large (98.45 %).

## Tables and figures

Table 1: Regressions on Placebo Variables

Estimated Effect of Neighborhood Diversity on:			
Fixed housing characteristics that should not be affected by $DIV_i$		Neighborhood amenities NOT under the control of HLM offices	
(1)		(2)	
Collective gas supply	-0.083 (0.091)	Streets maintenance	0.126 (0.130)
Ceiling height	0.014 (0.033)	Public transportation accessibility	0.077 (0.137)
Double glazing	-0.021 (0.088)	Private transportation accessibility	0.229* (0.136)
Recessed electrical installation	0.016 (0.095)	Green spaces	0.229 (0.157)
Electrical grounding	0.007 (0.032)		

Each coefficient is the estimated effect of diversity obtained from a separate regression using the baseline specification. Column (1) reports the coefficients for regressions of placebo variables which are fixed housing characteristics that should not be affected by  $DIV_i$ . Column (2) reports the coefficients for regressions of placebo variables which are neighborhood amenities not under the control of HLM offices. Standard errors clustered at the block level are reported in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.001$

<sup>16</sup>We replicate the regression of the dissatisfaction with housing conditions (Column 1) as well as of the three aggregate indices we obtained with the principal component analysis presented in section 1 of this online Appendix.

Table 2: New inhabitants' nationality and share of area population of same nationality

Dep Var: Share of population of the same ethnic group as new movers in a given area		
	(1)	(2)
Share of the <i>département</i> population of the same ethnic group		0.890*** (0.034)
<b>Nationality</b> (reference group: Other nationalities)		
Native	0.816*** (0.010)	0.067** (0.030)
Naturalized French	-0.044*** (0.012)	-0.037** (0.012)
European	-0.022* (0.012)	-0.007 (0.011)
Maghrebian	0.001 (0.014)	0.007 (0.013)
African	-0.025 (0.015)	-0.016 (0.015)
Asian	-0.085* (0.049)	-0.019 (0.047)
<b>Public Housing (HLM)</b>	-0.158** (0.068)	-0.184** (0.066)
<b>Nationality * HLM</b>		
HLM * Native	-0.038 (0.034)	0.040 (0.033)
HLM * Naturalized	0.095** (0.037)	0.051 (0.036)
HLM * European	-0.018 (0.038)	0.010 (0.037)
HLM * Maghrebian	0.055 (0.037)	0.024 (0.036)
HLM * African	0.046 (0.040)	0.017 (0.039)
HLM * Asian	0.083 (0.080)	0.027 (0.078)
<b>Household characteristics</b>		
Age	0.002** (0.001)	0.002** (0.001)
Age squared	-0.000** (0.000)	-0.000** (0.000)
Hourly wage (log)	-0.001 (0.002)	-0.001 (0.002)
Education	Yes	Yes
HLM * Education	Yes	Yes
SEC	Yes	Yes
HLM * SEC	Yes	Yes
<i>département</i> Fixed Effect	Yes	Yes
Intercept	0.085** (0.036)	0.070** (0.035)
R-squared	0.856	0.864
N	11519	11519

\* p&lt;0.10, \*\* p&lt;0.05, \*\*\* p&lt;0.001

Table 3: Rejection of HLM offers and Ethnic diversity

Rows: Dependent Variables	Coefficient associated with Ethnic Diversity			
	(1)	(2)	(3)	(4)
<b>Panel A:</b> Sample of households that currently live in public housing:				
I. Probability of having declined at least one HLM offer during the previous application process	0.058 (0.058)	0.069 (0.063)	0.017 (0.067)	0.123 (0.0886)
N	1,779	1,779	1,748	1,744
R <sup>2</sup>	0.001	0.021	0.023	0.089
II. Probability that the reason for having declined an HLM offer during the previous application was "unpleasant environment"	0.162 (0.144)	0.061 (0.158)	0.017 (0.171)	-0.0310 (0.258)
N	417	417	415	414
R <sup>2</sup>	0.003	0.035	0.050	0.308
<b>Panel B:</b> Sample of households that are currently applying to public housing:				
I. Probability of having declined at least one HLM offer during the current application process	-0.063 (0.057)	-0.043 (0.064)	-0.088 (0.071)	-0.116 (0.103)
N	1,192	1,192	1,173	1,171
R <sup>2</sup>	0.001	0.014	0.024	0.121
II. Probability that the reason for having declined an HLM offer during the current application was "unpleasant environment"	0.004 (0.194)	-0.007 (0.237)	-0.104 (0.250)	-0.122 (0.506)
N	198	198	195	194
R <sup>2</sup>	0.000	0.083	0.115	0.590

Each of the coefficients is estimated from a separate regression of each of the four dependent variables described in the first column on ethnic diversity. Column 1 does not include any control. Column 2 includes household characteristics (gender, age, education, employment status and nationality of the head of household, total income (in log) of the household per unit of consumption, and household size). Column 3 adds up the characteristics of the building (number of apartments (in log) and construction date). On top of that, column 4 includes neighborhood characteristics (socio-economic background (Tabard index), and local unemployment rate), as well as *département* fixed effects. In addition, a dummy variable indicating whether the household already lives in the public housing sector is included in specifications 2 to 4 of Panel B. Robust standard errors adjusted for block clustering are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Figure 1: Distributions of Diversity across Blocks



Table 4: Sensitivity to control variables

Control Variables:	Estimated Effect of Neighborhood Diversity:						
	No Control (1)	Building Age (2)	Tabard Index (3)	Unemployment Rate (4)	Household Size (5)	Household Income (6)	Duration of Tenancy (mean) (7)
<b>A. Neglect in the Common Areas:</b>							
Damages in the premises	1.443*** (0.142)	1.379*** (0.144)	1.245*** (0.154)	0.962*** (0.166)	1.367*** (0.145)	1.392*** (0.144)	1.453*** (0.141)
Graffiti	[1.165,1.721] 0.803***	[1.096,1.663] 0.752***	[0.943,1.548] 0.717***	[0.636,1.287] 0.608***	[1.084,1.651] 0.784***	[1.111,1.674] 0.775***	[1.177,1.728] 0.806***
Garbage on the floor	[0.663,0.942] 0.649***	[0.610,0.894] 0.601***	[0.568,0.867] 0.564***	[0.450,0.766] 0.459***	[0.644,0.924] 0.650***	[0.635,0.916] 0.645***	[0.668,0.944] 0.651***
Broken windows	[0.069] 0.513,0.784	[0.070] 0.464,0.739	[0.071] 0.426,0.703	[0.078] 0.305,0.612	[0.069] 0.515,0.785	[0.069] 0.509,0.781	[0.069] 0.517,0.786
Broken doors	[0.063] 0.394***	[0.066] 0.354***	[0.065] 0.328***	[0.068] 0.267***	[0.064] 0.379***	[0.063] 0.373***	[0.063] 0.396***
Broken light bulbs	[0.427,0.676] 0.559***	[0.390,0.647] 0.544***	[0.353,0.609] 0.523***	[0.318,0.584] 0.408***	[0.411,0.661] 0.537***	[0.420,0.666] 0.542***	[0.429,0.677] 0.560***
Broken mailboxes	[0.052] 0.456,0.662	[0.054] 0.439,0.650	[0.055] 0.416,0.630	[0.057] 0.297,0.519	[0.053] 0.433,0.640	[0.053] 0.439,0.645	[0.052] 0.458,0.663
Vandalism on the elevator	[0.641*** (0.068)	[0.608*** (0.070)	[0.568*** (0.071)	[0.472*** (0.073)	[0.623*** (0.068)	[0.618*** (0.068)	[0.644*** (0.067)
	[0.509,0.774] 0.319***	[0.471,0.745] 0.311***	[0.429,0.707] 0.282***	[0.328,0.616] 0.263***	[0.489,0.757] 0.314***	[0.485,0.751] 0.310***	[0.513,0.776] 0.321***
Noise in day time	[0.055] 0.210,0.427	[0.057] 0.200,0.423	[0.056] 0.171,0.392	[0.060] 0.145,0.381	[0.055] 0.206,0.423	[0.055] 0.202,0.418	[0.054] 0.214,0.428
	0.843*** (0.123)	0.685*** (0.126)	0.750*** (0.128)	0.464*** (0.137)	0.788*** (0.124)	0.770*** (0.125)	0.841*** (0.123)
Noise at night	[0.602,1.084] 0.806***	[0.437,0.932] 0.719***	[0.499,1.000] 0.746***	[0.196,0.732] 0.548***	[0.545,1.030] 0.770***	[0.525,1.014] 0.737***	[0.601,1.081] 0.806***
	[0.111] 0.589,1.023	[0.115] 0.493,0.945	[0.115] 0.520,0.972	[0.121] 0.311,0.786	[0.112] 0.551,0.989	[0.112] 0.518,0.956	[0.111] 0.589,1.024

Each coefficient is the estimated effect of diversity obtained from a separate regression. Each line corresponds to a different outcome, and each column includes a different control variable, as written in the first line. Each regression includes municipality fixed effects. Standard errors clustered at the block level are reported in parentheses. 95% confidence intervals are reported in brackets. Significance levels: \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

Table 4: Sensitivity to control variables

Control Variables:	Estimated Effect of Neighborhood Diversity:						
	No Control (1)	Building Age (2)	Tabard Index (3)	Unemployment Rate (4)	Household Size (5)	Household Income (6)	Duration of Tenancy (mean) (7)
<b>B. Poor Quality of Housing:</b>							
Care of the commons	1.011*** (0.128)	0.916*** (0.133)	0.898*** (0.139)	0.453** (0.153)	0.924*** (0.128)	0.994*** (0.129)	1.011*** (0.128)
Cold in apartment	[0.760,1.261] 0.406*** (0.065)	[0.655,1.176] 0.365*** (0.068)	[0.625,1.170] 0.371*** (0.067)	[0.154,0.752] 0.285*** (0.076)	[0.672,1.176] 0.384*** (0.065)	[0.741,1.248] 0.370*** (0.066)	[0.760,1.263] 0.406*** (0.065)
Cold due to bad insulation	[0.278,0.534] 0.096** (0.045)	[0.231,0.499] 0.069 (0.047)	[0.241,0.502] 0.090** (0.045)	[0.137,0.434] 0.059 (0.049)	[0.257,0.512] 0.084* (0.045)	[0.240,0.500] 0.083* (0.046)	[0.278,0.534] 0.096** (0.045)
Cold due to equipment breakdown	[0.009,0.183] 0.051 (0.033)	[-0.023,0.161] 0.052 (0.034)	[0.003,0.178] 0.037 (0.034)	[-0.038,0.155] 0.023 (0.039)	[-0.004,0.173] 0.048 (0.033)	[-0.006,0.173] 0.052 (0.033)	[0.008,0.183] 0.052 (0.033)
Cold due to poor equipment	[-0.013,0.116] 0.216*** (0.044)	[-0.015,0.118] 0.191*** (0.045)	[-0.029,0.104] 0.189*** (0.045)	[-0.054,0.100] 0.148** (0.053)	[-0.017,0.113] 0.205*** (0.043)	[-0.014,0.117] 0.190*** (0.044)	[-0.013,0.116] 0.216*** (0.044)
Soundproofing quality	[0.131,0.302] 0.942*** (0.139)	[0.102,0.280] 0.552*** (0.139)	[0.101,0.277] 0.863*** (0.147)	[0.045,0.251] 0.413** (0.160)	[0.121,0.290] 0.880*** (0.140)	[0.104,0.276] 0.870*** (0.141)	[0.130,0.302] 0.933*** (0.137)
Elevator breakdown	[0.669,1.215] 0.334*** (0.066)	[0.279,0.825] 0.327*** (0.067)	[0.575,1.151] 0.288*** (0.065)	[0.100,0.727] 0.317*** (0.069)	[0.605,1.155] 0.326*** (0.066)	[0.599,1.153] 0.336*** (0.066)	[0.664,1.201] 0.337*** (0.065)
Toilet malfunction	[0.205,0.463] 0.229*** (0.057)	[0.195,0.459] 0.174** (0.060)	[0.160,0.415] 0.203*** (0.059)	[0.181,0.453] 0.130** (0.065)	[0.198,0.455] 0.197*** (0.057)	[0.207,0.466] 0.190** (0.058)	[0.209,0.465] 0.229*** (0.057)
<b>C. Public Safety:</b>							
Robberies	[0.117,0.341] 0.055 (0.042)	[0.057,0.291] 0.063 (0.044)	[0.086,0.319] 0.048 (0.043)	[0.003,0.257] 0.075 (0.047)	[0.085,0.308] 0.042 (0.043)	[0.077,0.304] 0.049 (0.043)	[0.116,0.341] 0.056 (0.042)
Aggressions	[-0.028,0.138] 0.115** (0.041)	[-0.024,0.150] 0.116** (0.043)	[-0.036,0.133] 0.090** (0.044)	[-0.017,0.167] 0.036 (0.047)	[-0.041,0.125] 0.097** (0.042)	[-0.035,0.132] 0.097** (0.042)	[-0.027,0.138] 0.116** (0.041)
Burglary	[0.034,0.196] 0.042 (0.028)	[0.032,0.200] 0.031 (0.029)	[0.003,0.178] 0.028 (0.031)	[-0.057,0.128] -0.002 (0.030)	[0.015,0.178] 0.044 (0.028)	[0.015,0.179] 0.033 (0.028)	[0.036,0.197] 0.042 (0.028)
	[-0.013,0.096] 0.042 (0.028)	[-0.025,0.087] 0.031 (0.029)	[-0.032,0.088] 0.031 (0.031)	[-0.061,0.058] -0.002 (0.030)	[-0.011,0.099] 0.044 (0.028)	[-0.023,0.089] 0.033 (0.028)	[-0.012,0.096] 0.042 (0.028)

Each coefficient is the estimated effect of diversity obtained from a separate regression. Each line corresponds to a different outcome, and each column includes a different control variable, as written in the first line. Each regression includes municipality fixed effects. Standard errors clustered at the block level are reported in parentheses. 95% confidence intervals are reported in brackets. Significance levels: \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

Table 5: Ethnic Diversity and Dissatisfaction with Housing conditions – Robustness Checks  
(to be continued)

	Dissatisfaction with Housing Conditions		
	(1)	(2)	(3)
Ethnic Diversity	1.392** (0.530)		
<b>Household characteristics:</b>			
Gender	-0.018 (0.032)	-0.019 (0.032)	-0.021 (0.032)
Age	-0.005*** (0.001)	-0.005*** (0.001)	-0.005*** (0.001)
Education	-0.013* (0.007)	-0.014* (0.007)	-0.014* (0.007)
Income (log)	-0.041 (0.031)	-0.041 (0.031)	-0.040 (0.031)
Household size	0.105*** (0.013)	0.104*** (0.013)	0.105*** (0.013)
<i>Employment status (ref: Employed)</i>			
Unemployed	0.050 (0.055)	0.059 (0.055)	0.051 (0.055)
Inactive	-0.050 (0.047)	-0.049 (0.047)	-0.051 (0.047)
<i>Nationality (ref: French at birth)</i>			
Naturalized French	-0.044 (0.066)	-0.043 (0.066)	-0.036 (0.066)
European	0.059 (0.088)	0.058 (0.089)	0.062 (0.089)
Maghrebian	-0.107 (0.068)	-0.106 (0.068)	-0.108 (0.068)
African	0.092 (0.143)	0.098 (0.143)	0.097 (0.143)
Asian	-0.615* (0.316)	-0.607* (0.318)	-0.605* (0.315)
Other nationality	0.577 (0.630)	0.565 (0.632)	0.569 (0.634)
<b>Building characteristics:</b>			
Nb of dwellings (log)	0.017 (0.013)	0.020 (0.013)	0.019 (0.013)
<i>Date of construction of the building (ref: before 1948)</i>			
1949 ≤ t < 1974	0.007 (0.077)	0.022 (0.077)	0.012 (0.076)
1975 ≤ t < 1981	-0.115 (0.082)	-0.101 (0.081)	-0.110 (0.081)
1982 ≤ t < 1989	-0.134 (0.082)	-0.122 (0.082)	-0.129 (0.082)
1990 ≤ t < 1998	-0.444*** (0.085)	-0.438*** (0.086)	-0.439*** (0.085)
1999 ≤ t	-0.769*** (0.168)	-0.770*** (0.168)	-0.751*** (0.170)

Robust standard errors adjusted for block clustering are in parentheses.  
\*\*\*p<0.01, \*\* p<0.05, \* p<0.1

Table 5: Continued

Dependent Variable:	<b>Dissatisfaction with housing condition</b>		
	(1)	(2)	(3)
<b>Neighborhood characteristics:</b>			
Block unemployment rate	0.909*** (0.175)	0.954*** (0.175)	0.911*** (0.174)
Socio-economic background (Tabard)	Yes	Yes	Yes
<i>Département</i>	Yes	Yes	Yes
Fixed Effects			
<b><i>Ethnic group shares:</i></b>			
% European	-3.249** (1.007)	-0.928* (0.493)	-2.273** (0.961)
% Maghrebien	-1.092 (0.707)	0.633** (0.277)	2.014*** (0.521)
% African	-1.203 (1.101)	0.941 (0.823)	1.218 (1.428)
% Asian	-2.186* (1.323)	0.167 (0.974)	0.315 (1.838)
% Other nationality	-1.488* (0.855)	0.260 (0.501)	-0.263 (0.864)
<b><i>Squared ethnic group shares:</i></b>			
(% European) <sup>2</sup>			7.587* (4.332)
(% Maghrebien) <sup>2</sup>			-3.767** (1.220)
(% African) <sup>2</sup>			-4.225 (7.839)
(% Asian) <sup>2</sup>			-4.792 (14.523)
(% Other nationality) <sup>2</sup>			1.351 (2.895)
Intercept	4.009*** (0.361)	4.071*** (0.362)	4.023*** (0.363)
R-squared	0.131	0.129	0.131
Observations	4379	4379	4379

Robust standard errors adjusted for block clustering are in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 6: A proxy for language fractionalization

Dependent Variable:	Dissatisfaction with housing conditions (1)	Neglect of the commons (2)	Quality of housing (3)	Insecurity (4)
<i>Panel A:</i>				
Diversity based on nationality at birth	0.368*** (0.129)	1.532*** (0.422)	0.727*** (0.263)	0.0252 (0.183)
Observations	4,379	1,693	3,797	4,379
R-squared	0.156	0.201	0.192	0.063
<i>Panel B:</i>				
Diversity based on share of the population speaking French in country of origin	0.366*** (0.141)	1.560*** (0.472)	0.741** (0.290)	-0.0377 (0.199)
Observations	4,365	1,689	3,788	4,365
R-squared	0.157	0.202	0.193	0.064

Standard errors in parentheses \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. Controls: head of household characteristics (gender, age, education, activity status, aggregated nationality), household characteristics (log of income, household size), building characteristics (date of construction, log number of housing units), socio-economic background of the neighborhood (unemployment rate, Nicole Tabbarid classification), *département* fixed-effects. Each coefficient comes from a separate regression. The columns indicate the four dependent variables under study. Panel A and panel B respectively correspond to the measure of diversity used in each regression.

## References

- Fougère, D., F. Kramarz, and J. Pouget (2009, 09). Youth unemployment and crime in France. *Journal of the European Economic Association* 7(5), 909–938.
- Goux, D. and E. Maurin (2007, October). Close neighbours matter: Neighbourhood effects on early performance at school. *Economic Journal* 117(523), 1193–1215.
- Hémet, C. (2013, October). The local determinants of victimization. AMSE Working Papers 1349, Aix-Marseille School of Economics, Marseille, France.
- Vigdor, J. L. (2002, April). Interpreting ethnic fragmentation effects. *Economics Letters* 75(2), 271–276.

**D.2 Diversity and Employment Prospects: Neighbors Matter!**

---

# Diversity and Employment Prospects Neighbors Matter!

---

**Camille Hémet**  
**Clément Malgouyres**

## ABSTRACT

*Using recent data from the French Labor Force Survey, this paper explores how diversity affects individuals' employment prospects at various geographic levels. Employment correlates positively with local labor market diversity, but negatively with neighborhood diversity. Using several approaches to deal with the endogeneity of local labor market diversity, we do not find any robust evidence of a causal impact of diversity on employment at this rather aggregate level, suggesting that immigrants actually tend to self-select into more economically dynamic areas. However, taking advantage of the very precise localization of the data in order to correct biases related to residential sorting, we confirm a negative effect of neighborhood diversity. We also show that diversity in terms of nationalities matters more than diversity based on parents' origins, giving insights on the underlying mechanisms.*

---

*Camille Hémet is assistant professor at École Normale Supérieure and Paris School of Economics and a research fellow with the CEPR. Clément Malgouyres is a research economist at Banque de France. We are grateful to Yann Algan, Bruno Decreuse, Nina Guyon, Jordi Jofre Monseny, Alan Manning, Thierry Mayer, Matthieu Solignac, Maxime To, Javier Vásquez Grenno, and Yves Zenou for insightful comments and discussion. We are deeply indebted to two anonymous referees whose comments helped us substantially improve the paper. We also thank participants at the following events for their remarks: University of Strasbourg Cournot seminar, University of Lund seminar, 10th UEA meeting, 2014 SOLE meeting, IEB seminar, National University of Singapore seminar, PSE Applied Economics lunch seminar, 62nd annual meeting of the French Economic Association (AFSE) and WZB workshop on Ethnic Diversity. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'avenir" program (reference: ANR-10-EQPX-17, Centre d'accès sécurisé aux données, CASD) and by two other public grants overseen by the ANR (references: Labex OSE ANR-10-LABX-93-01 and ANR-16-FASI-0001-01). The views expressed here are those of the authors only and do not necessarily reflect the views of the Banque de France or the Eurosystem. This paper uses data from the labour force survey produced by the French national statistical institute (INSEE) that are sensitive because they are localized at the neighborhood level. The data can be obtained through the Secure Remote Access Center (CASD, <https://www.casd.eu/en/>) after filing a request to the French Committee for the Statistical Secret (procedure explained here, in French, [http://www.cnis.fr/cms/Accueil/activites/\\_trois\\_comites/Comite\\_du\\_secret\\_statistique/Les\\_procedures\\_dacces\\_aux\\_donnees\\_confidentielles](http://www.cnis.fr/cms/Accueil/activites/_trois_comites/Comite_du_secret_statistique/Les_procedures_dacces_aux_donnees_confidentielles)). The authors are willing to assist (Camille Hémet, [camille.hemet@psemail.eu](mailto:camille.hemet@psemail.eu)).*

[Submitted January 2015; accepted April 2017]; doi:10.3368/jhr.53.3.0115-6895R1

JEL Classification: J15, J60, R23, and Z13

ISSN 0022-166X E-ISSN 1548-8004 © 2018 by the Board of Regents of the University of Wisconsin System

## I. Introduction

Western economies are facing intensified flows of immigration due to trade development and economic integration and consequently have to cope with increasingly mixed populations. Following the recent European Union enlargement process and the current massive immigrant inflows, this feature is particularly salient in European Union countries, where the economic and social implications of higher heterogeneity are becoming central issues. In particular, public opinion is generally hostile to immigrants, who are often perceived as a threat to job security and wages, although this fact is not clearly supported by empirical research. In a recent influential paper on the topic, Ottaviano and Peri (2012) adopt a general equilibrium approach and show that the massive immigration to the United States over the 1990–2004 period actually increased natives' wages, contradicting the seminal work by Borjas (2003). In contrast to the large literature studying the economic impact of immigration on natives, papers looking at the labor market effect of diversity per se are scarce. Using U.S. cities data, Ottaviano and Peri (2006) jointly estimate wage and rent equations and find that diversity is positively associated with both variables. They conclude that diversity has a net positive impact on U.S.-born workers' productivity. Using a similar setting, Bellini et al. (2013) replicate these results for European regions. To the best of our knowledge, these are the only two papers in the diversity literature dealing with labor market outcomes, although they focus on wages rather than employment.

This paper intends to fill this gap by assessing the impact of local diversity on individuals' employment prospects. It asks the following question: To what extent do people living in heterogeneous neighborhoods have different employment probabilities than those living in more homogeneous areas? In other words, this work investigates how individuals cope with increasing levels of diversity, and in particular, how this affects their employment prospects. It is relevant in the current context of high unemployment, especially in high immigration countries such as France, Italy, and more recently, Spain. At the micro level, if different ethnic or cultural groups are hermetic to each other, in the sense that no interaction takes place across groups, then diversity can act as a barrier to communication, and in particular, to job information transmission. Given the importance of personal networks in the job search process (see Ioannides and Datcher Loury 2004), diversity could therefore reduce the chance of finding a job. On the other hand, if communication across groups is not an issue, then mixing people conveying nonredundant pieces of information (for example, due to different backgrounds) can certainly improve employment prospects. At a more aggregate level, diversity can affect employment probability through its impact on productivity, leading to ambiguous effects as well. On the bright side, diversity can be beneficial to productivity due to complementarity in workers' skills (see Lazear 1999; Alesina and La Ferrara 2005). On the downside, heterogeneity can hinder productivity by preventing social capital formation (Coleman 1988).<sup>1</sup>

---

1. Not only is the impact of diversity on productivity unclear, but the impact of productivity on employment is ambiguous as well: Nordhaus (2005) finds that more rapid productivity growth leads to increased rather than decreased employment in manufacturing, a sector that recently experienced a large employment decline. On the contrary, Michelis et al. (2013) find a strong negative relationship between TFP growth and labor input.

As we see from this brief discussion, the effect of diversity on employment is theoretically ambiguous. We investigate this question using detailed French employment data that allow us to disentangle the neighborhood (local) and labor market (aggregate) effects of diversity. We consider this exercise as the main contribution of our work. Specifically, we measure diversity using several definitions of origins and at various geographic levels, so as to understand as precisely as possible the mechanisms underlying the diversity–employment relationship. In addition, we adopt several identification strategies in order to bypass the endogeneity issue that is likely to flaw any estimate of the impact of diversity. These three methodological elements, which allow us to answer the central question of this paper, are detailed below.

First, the level of diversity is measured at two different geographic levels. As discussed above, diversity could impact employment prospects locally through networks and on a larger scale through productivity. In order to account for both effects, we compute diversity at a very local neighborhood level and at the employment zone (local labor market) level. As far as we know, it is the first time that various geographic scales for diversity are simultaneously considered. In addition, this and Algan et al. (2016) are the first studies conducted at such disaggregated levels. The results reveal that employment probability is negatively correlated with neighborhood diversity, but positively correlated with employment zone diversity, suggesting a negative effect through networks and a positive one through productivity.

Second, we rely on three alternative definitions of origins to measure diversity, namely nationality, birth country, and parents' origins, while the existing literature mostly relies on ethnolinguistic and religious groups. The measure based on parents' origins encompasses first- and second-generation immigrants and is therefore more closely related to the standard ethnic classification of individuals. By contrast, defining kinship according to nationality introduces the notion of immigrants' integration through naturalization. This distinction allows us to draw conclusions on whether the cultural or the ethnic dimension of diversity prevails. An important finding of this paper is that diversity based on nationality has a larger impact than diversity based on birth country, which is itself more relevant than diversity based on parents' origins, suggesting a preeminent role of cultural over ethnic diversity.

Third, we tackle the endogeneity issue that is pervasive in the literature on ethnic diversity. An important concern is that individuals have a preference for living close to their co-ethnics and thus tend to gather along ethnic lines, biasing any measure of the effect of diversity. Another issue is that of reverse causality that can arise if immigrants decide to settle in more economically dynamic areas. The endogeneity of employment zone diversity is handled through a traditional instrumental variable approach, where two different instruments are proposed. Following Card (2001) and Saiz (2007), we construct the predicted level of diversity in each employment zone based on the distribution of each origin group across employment zones in 1968 and the current number of individuals from each origin in France at the time of the study (2007–2013). An alternative and more innovative instrument is the level of diversity within the public housing tenants of the employment zone. It builds on Algan et al. (2016), who show that the allocation of households across public housing units in France does not take their origins or their preference for diversity into account, so that public housing diversity can be considered as exogenous. None of the instrumental variable (IV) estimates are

statistically significant and have large standard errors—despite strong first stages. The imprecision of the estimate hence precludes any firm conclusion regarding the absence of a causal effect of diversity at the aggregate level. However, additional results obtained using the specific subsample of public housing tenants, for which diversity can be considered as exogenous, point toward a null effect of aggregate diversity and tend to support the intuition that the positive aggregate relationship suggested by ordinary least squares (OLS) estimates was actually driven by selection. To deal with the endogeneity of local neighborhoods diversity, we follow Bayer et al. (2008) and assume that although households are able to select the precise area in which they want to live, they are, however, unable to pinpoint an exact neighborhood within this given area. Therefore, after controlling for sorting in a restricted area encompassing local neighborhoods, the assignment of individuals to a specific neighborhood within the selected area is considered as essentially random and provides a credibly exogenous source of variation to identify the effect of diversity. After controlling for residential sorting, the effect of neighborhood diversity remains significantly negative, though slightly attenuated. This suggests that sorting resulted in more economically deprived individuals living in more diverse neighborhoods.<sup>2</sup>

The remainder of the paper is organized as follows. In Section II, we extensively discuss the various channels through which diversity can affect employment prospects, building on existing empirical and theoretical studies. We carefully document our data set and its particular sampling design in Section III, where we also describe the various measures of diversity. Section IV deals with the main analysis of the paper. After presenting our baseline specification in Section IV.A, we deal with the issue of endogeneity in Sections IV.B and IV.C. In Section V, we propose an interpretation of the previous results and their underlying mechanisms, before concluding in Section VI.

## II. Diversity and Employment Prospects

The interest in the effect of diversity on economic performance and social peace has been rooted in economic research since the seminal paper by Easterly and Levine (1997) showing that Africa's high levels of ethnic diversity help explain its "tragic growth performance." The subsequent literature covers a very broad set of issues. Diversity is generally found to reduce public good provision because the threat of sanction to punish defectors is not credible across groups or because different groups do not share the same preferences and cannot agree on the type of public good to be produced.

---

2. Given theoretical mechanisms suggesting that diversity affects labor market outcomes through the productivity channel, investigating the wage effect of diversity could be informative. However, any analysis of the wage effect of diversity requires circumventing the issue of selection into employment, particularly in a high-unemployment country such as France. Finding a credible instrument explaining employment but not affecting wages is notoriously difficult, mainly because the exclusion restriction is delicate to justify. Moreover, measurement error in hourly wages in the French LFS is a serious issue (see for example, Cahuc and Carcillo 2014). With this caveat in mind, we conducted the analysis on wages without taking selection into account. In all OLS specifications, coefficients have the same sign than in the employment equations. The IV regressions lead to somewhat different results as the significantly positive correlation at the aggregate level remains. We believe that these results are hard to interpret, especially as the sign of the bias associated with selection is not clear. These results are not reported in the paper for the reasons mentioned above, but are available upon request.

This result holds in developing countries and developed countries alike—see Miguel and Gugerty (2005) for Kenya, Alesina et al. (1999) for the United States, and Algan et al. (2016) for France.<sup>3</sup> Another trend of the literature focuses rather on the social impact of diversity and shows that it is associated with lower participation in civic life or community activities (Alesina and La Ferrara 2000; Costa and Kahn 2003) and reduced trust (Alesina and La Ferrara 2002).

The present paper focuses on employment and is therefore more closely related to the branch of the literature that studies productivity. At the macro level, diversity can affect employment through its effect on productivity. A large part of the literature supports the idea that diversity has a positive impact on diversity related to skills complementarity, which dominates the negative effects linked to coordination issues. Indeed, workers from different origins are more likely to have been exposed to diverse cultures and distinct school systems (especially if they come from different countries), acquiring various skills and learning different approaches to the same problem, so that their collaboration can increase productivity and facilitate innovation. More formally, Hong and Page (2001) develop a model showing that teamwork may benefit more from low-skill but cognitively diverse workers than from homogenous high-skill workers. In a different theoretical setting, Lazear (1999) shows that when multicultural workers are complementary in the sense that they can exchange nonredundant and relevant information, the benefits from diversity offset its costs (for example, barriers to communication).

Several recent papers also bring empirical support to the beneficial impact of diversity on productivity and economic performance more generally. Using data from 160 metropolitan areas in the United States, Ottaviano and Peri (2006) jointly estimate wage and rent equations and find that diversity, measured in terms of birth countries, is positively associated with both variables. These results are robust to the inclusion of many confounding factors proxying for productivity and amenity shocks across cities, as well as to the instrumentation of diversity to correct for endogeneity. They conclude that diversity has a net positive impact on U.S. born workers' productivity. A similar methodology is adopted by Bellini et al. (2013), who reach the same conclusion for European regions. Finally, Alesina et al. (2016) investigate the relationship between birth country-based diversity and economic development in a cross section of countries. Potential endogeneity due to reverse causality is addressed through instrumental variable estimation. They compute a predicted measure of immigrants' diversity by estimating a gravity model based on exogenous geographic and cultural bilateral variables. They find that while standard ethnolinguistic fractionalization is detrimental to economic success, the impact of diversity in terms of birth countries is positive, especially in more developed countries.<sup>4</sup>

At a more micro level, diversity may affect individuals' employment prospects through the channel of networks and job information transmission, for instance, if it prevents communication about relevant information. There is considerable evidence that information transmission plays a key role in the labor market.<sup>5</sup> Many empirical

3. An exception is Glennerster et al. (2013), who do not find any particular effect of diversity in Sierra Leone villages.

4. This paper also provides a comprehensive review of the literature on the costs and benefits of diversity.

5. Ioannides and Datcher Loury (2004) and Ioannides and Topa (2010) provide comprehensive surveys on the topic.

studies conducted over various time periods and in diverse countries agree that relying on friends and family is a very popular job search method and that, on average, half of jobs are found through social networks (Corcoran et al. 1980; Granovetter 1995; Holzer 1988; Wahba and Zenou 2005). Theoretically as well, Calvó-Armengol and Jackson (2004) show that employment probability increases both with the number of links an agent has and with the employment rate in the individual's network. In particular, several papers focus on the role of ethnic and immigrant networks. A recent paper by Battu et al. (2011) shows that ethnic minorities in the UK rely extensively on personal networks when searching for a job, although this does not necessarily lead to better employment prospects. The sociology literature also emphasizes the importance of ethnic networks in business relations and entrepreneurship, through an increased capacity to cooperate due to common languages and values (Light and Rosenstein 1995; Light 2005).

On the other hand, a few recent works investigating the racial dimension of social networks in the United States provide evidence consistent with the notion that interactions within racial groups is more intense than across. This would tend to support the idea that diversity could hinder communication, hence reducing employment prospects. Although this literature remains limited, Hellerstein et al. (2008) show for instance that the employment probability of black workers is primarily affected by black job density rather than job density as a whole. Ananat et al. (2013) also demonstrate that black workers' wages benefit less from agglomeration economies than white workers' wages. The race specificity of gains associated with higher levels of economic density suggests that the strength of agglomeration economies is weaker across than within races, presumably because of a lower level of knowledge spillovers and exchange of ideas between workers in the same location (a demonstrated channel underpinning agglomeration economies). Some insight can also be gained from the literature on "oppositional identities," which investigates the fact that individuals may decide to reject the dominant culture. In particular, the size or the racial composition of a community are put forward as potential determinants of such behavior (see, for instance, Bisin et al. 2011). Thus, individuals living in more diverse communities may tend to develop oppositional identities (depending on the relative size of the different groups for instance), which could in turn affect their ability to search for jobs (either due to a different access to a given social network or possibly due to different social skills and norms than those of the dominant group).

Because communication across ethnic groups may be hindered by a tendency to self-segregate by different religious beliefs and culture and, above all, by differences in the languages spoken, diversity may prevent network formation and information transmission, thus having a negative impact on individuals' labor market performances. This effect could be amplified if diversity exists at the neighborhood level, as networks tend to be very local (see for instance Wellman, 1996). In particular, a few recent studies have shown that local social interactions within neighborhoods do affect employment and wage outcomes. For instance, Weinberg et al. (2004) show that a one standard deviation increase in neighborhood employment is associated with a 6.1 percent increase in annual hours worked for adult males on average. Bayer et al. (2008) estimate that living in the same block increases by more than 33 percent the probability to work at the same location. In a paper dealing explicitly with ethnic networks, Patacchini and Zenou

(2012) show that the individual probability of finding a job increases with the number of ties, but that the magnitude of the effect decreases with distance. To summarize, if individuals are unable to create social ties within their neighborhood because they live in a diverse environment, this might hinder their ability to search and find jobs through the network.

### III. Data and Descriptive Statistics

This paper relies on the French Labor Force Survey (*Enquête Emploi*, INSEE, hereafter the LFS), a quasipanel surveying about 72,000 individuals quarterly since 2003. The LFS is characterized by a particular sampling scheme that makes it extremely valuable for studying neighborhood effects. Because it is central to our identification of neighborhood diversity effects, we start by providing a detailed description of this sampling design. The French territory is initially partitioned into about 2,500 “primary units,” which are in turn divided into “sectors.” Each sector is demarcated by obvious topographical boundaries, such as streets, roads, railways, or rivers, and is designed to include between 120 and 240 homes. One sector is then randomly selected in each primary unit, with a probability proportional to the number of homes it contains. The baseline sample is thus made of about 2,500 sectors. These sectors are in turn split into contiguous areas of 20 homes on average, corresponding to what we refer to as “local neighborhoods” (*aires* in French). Sectors containing 120 homes are divided into exactly six adjacent local neighborhoods, while larger sectors (a minority) may include more than six local neighborhoods. In the former case, all six local neighborhoods are included in the baseline sample; otherwise, six local neighborhoods are drawn randomly. The baseline sample eventually contains all individuals living in the six contiguous local neighborhoods included in each sector. Thus, the sample is not drawn from a selection of homes, but from a selection of local neighborhoods. Only a fraction of this baseline sample is then included in the rotating panel for each wave. More specifically, the first neighborhood of a sector is exhaustively surveyed over six quarters, before being replaced by the second one for the next six quarters. This process goes on until all six neighborhoods of the sector are surveyed.<sup>6</sup> This particular sampling strategy implies that we observe *all* individuals living in these local neighborhoods, so that we can perfectly characterize their socioeconomic and demographic environment. Interestingly, it also implies that we are able to observe individuals living in contiguous local neighborhoods within the same sector, a feature that is key to our identification of neighborhood diversity effects, as will be explained in Section IV.C.<sup>7</sup>

The LFS contains precise information about individuals’ labor market situation: employment status, wage, type of contract, tenure, job search methods, and socioeconomic category. It also provides relevant individual characteristics, such as age, gender, education, and marital status. Because we are interested in individuals’ employment status, we restrict the sample to working-age individuals (16–65 years), and we drop the

6. Note that not all sectors were included at the same time in the main sample so that not all households are replaced at once. In fact, only 1/6 of the sample is renewed each quarter.

7. Refer to INSEE documentation for more details on the sample composition and selection.

**Table 1**  
*Individual Characteristics*

	Min.	Max.	Mean	(SD)	Median
<b>Demographics</b>					
Age	16	65	44.2	(12.4)	45
Male	0	1	49		
<b>Employment Status</b>					
Employed	0	1	0.718		
Inactive	0	1	0.282		
<b>Socioeconomic category</b>					
Craftsman, shopkeeper	0	1	0.059		
Executive or other high position	0	1	0.146		
Intermediate occupation	0	1	0.231		
Employee	0	1	0.307		
Factory worker	0	1	0.244		
Unemployed never employed	0	1	0.011		
<b>Level of education</b>					
No degree	0	1	0.20		
Degree lower than high school	0	1	0.343		
High school	0	1	0.174		
College	0	1	0.283		

Notes:  $N=396,823$ . The sample contains the first and last observations of each sampled individual during years 2007–2013. The sample is described in more detail in Section III.

students. Importantly, we can infer individuals' origins using three alternative variables: nationality, birth country, and parents' nationality at birth.

This last variable was not properly included in the survey before 2007, so we limit the study to 2007–2013. A final sample restriction comes from the needs of the strategy used to identify the effects of local neighborhood diversity. The general idea is to compare individuals surveyed in a given neighborhood during a given quarter to individuals surveyed in an adjacent neighborhood during the subsequent quarter. This requires us to keep only the first and the first and the last (sixth) observations of each individual, as is carefully explained in Section IV.C. In the end, our sample contains about 396,000 observations. Table 1 summarizes the main employment-related individual characteristics for this subsample.

We characterize individuals' origins in three different ways (Table 2). Not only do we use their nationality and country of birth as reported in the survey, but we also build a third measure combining information on individuals' origins and their parents'. In contrast to the rest of the literature, which mostly relies on a single measure of origin (ethnolinguistic group), we argue that these alternative measures of origins capture different individual traits and characteristics, as explained in the following discussion.

**Table 2**  
*Nationality, Country of Birth, and Parents' Origin Share by Group of Countries*

Group of Countries	Nationality	Country of Birth	Parents' Origin
French	0.945	0.883	0.813
Southern Europe	0.015	0.021	0.062
Rest of Europe	0.010	0.018	0.034
North Africa	0.014	0.041	0.052
Rest of Africa	0.007	0.016	0.016
Row	0.009	0.020	0.024

Notes:  $N=396, 823$ . The sample contains the first and last observations of each sampled individual during years 2007–2013. The sample is described in more detail in Section III.

We define the “parents’ origins” group such that an individual belongs to a given non-French category if at least one of her or his parents belongs to this group. For instance, a French person, born in France, but whose parents were born Tunisian, is allocated to the North African (or Maghreb) group. This measure of origins presents the advantage of taking first and second generation immigrants into account: it enables us to impute a foreign origin to an individual, even when she was born French in France. More importantly, this measure of origins is the most closely related to the standard ethnic classification of individuals,<sup>8</sup> while measures of nationality or birth country would less precisely reflect an individual’s ethnicity. For instance, a person whose parents were born Senegalese is very likely to be black, even though she is French and born in France. By contrast, defining kinship according to nationality introduces the notion of immigrants’ integration through naturalization. Indeed, the nationality of foreign-born individuals living in France usually corresponds to their birth country, unless they have become French through a naturalization process. We therefore argue that measuring origins based on nationality gets closer to reflecting cultural rather than ethnic diversity. Indeed, two individuals sharing the same nationality are more likely to speak a common language and to share other cultural traits. Obviously, this can be true for two native individuals from the same country, who share a set of cultural norms from this country. But, more importantly, this also holds for two individuals who share the same nationality even if one of them is naturalized. Indeed, in the case of France, acquiring the French nationality implies having spent several years in France, learning French, getting used to the country’s bureaucracy, its social norms, and so on. Yet, we do not claim that two individuals sharing the same nationality (for example, French) but coming from different backgrounds (for example, different birth countries or different parental origins) share exactly the same set of cultural norms and social codes because they would have inherited different ones from their different backgrounds. Rather, what we argue is that these two individuals *do share a common language and set of norms* (that is, the French

8. At least, it is the best proxy for ethnic origins we can get using French data where ethnic classifications are forbidden by law.

ones), which they acquired in the process of becoming French, *on top of the different norms they inherited from their family background*. To summarize, we think that an individual's origin as defined by her parents' reflects her ethnic background much more precisely than what nationality could do, while her nationality is a mark of a cultural belonging that comes in addition to other possible cultural traits inherited from different origins (for example, from birth country and parents).

Using these various classifications of origins, we compute three corresponding measures of diversity. The level of diversity in a given area reflects the probability that two randomly drawn individuals from the population belong to two different groups. Formally, it is computed using the standard fractionalization index used in the literature (see Alesina et al. 2003 for an extensive description):

$$(1) \quad DIV_j = \sum_{i=1}^j s_{ij}(1 - s_{ij}) = 1 - \sum_{i=1}^j s_{ij}^2$$

where  $s_{ij}$  is the share of individuals from group  $i$  ( $i = 1, \dots, 6$ ) (as defined above) in geographic area  $j$ . This index takes its minimum at 0 when the population living in area  $j$  is fully homogeneous, and it converges to 1 as the population heterogeneity increases. As argued in Section II, diversity may impact employment through different channels at different geographical areas. To test for this, we assess diversity at two different levels. The first level, called "employment zone," corresponds to a local labor market. It is defined as an area within which most of the labor force lives and works, and in which firms can find most of the labor force necessary to fill the vacant positions. There are about 300 employment zones in mainland France. Measuring diversity at an aggregate level corresponding to a consistent local labor market is particularly useful to capture the potential effect of diversity on employment through the channel of productivity. The second level is much more disaggregated and corresponds to the local neighborhood made of about 20 households (LFS sampling unit). Measuring diversity at such a local level aims at accounting for local interactions between immediate neighbors to test the job information transmission mechanism. In addition, using three different definitions of origins enables us to capture various dimensions of diversity. Following the previous discussion, we argue that parents' origins-based diversity is the best proxy for ethnic diversity, while diversity based on nationality tends to reflect cultural diversity (for example, diversity of language, social norms, or cultural habits) rather than ethnic diversity. Finally, we consider birth country diversity as an intermediate measure.

Table 3 describes the levels of diversity prevailing in individuals' local neighborhoods and employment zones. It is immediate to see that diversity is the lowest when measured in terms of nationalities and the highest when computed based on individuals' and parents' origins, which is not surprising. Consider for instance an area made of three French individuals, one born in France from French parents (for example, native French), one born in France from Vietnamese parents (for example, second generation immigrant), and one born in Morocco from Moroccan parents (for example, first generation immigrant). This population is completely homogeneous ( $DIV_j = 0$ ) if we consider the individuals' nationality only. However, diversity is larger (0.44) once birth countries are taken into account, and even more (0.67) once parents' origins is considered. Note also that employment zone diversity is slightly larger than neighborhood diversity.

**Table 3**  
*Descriptive Statistics Regarding the Diversity Index*

Statistics	Nationality	Birth Country	Parents' Origin
Local neighborhood level			
Mean	0.0961	0.1891	0.2742
Standard deviation	0.1429	0.1838	0.2065
Median	0	0.1420	0.2400
Employment zone level			
Mean	0.1140	0.2164	0.3199
Standard deviation	0.0925	0.1392	0.1694
Median	0.0879	0.1839	0.3025

Notes:  $N=396, 823$ . The sample contains the first and last observations of each sampled individual during years 2007–2013. The sample is described in more detail in Section III.

## IV. Results

### A. Baseline Specification

In this section, diversity is considered as exogenous, and its impact on employment status is estimated with the following equation:

$$(2) \quad EMP_{it} = \alpha + \beta_1 DIV_{z(i)t} + \beta_2 DIV_{n(i)t} + \eta_1 W_{z(i)t}^z + \eta_2 W_{n(i)t}^n + \gamma X_{it} + \varphi_t + \varepsilon_{it}$$

where  $EMP_{it}$  is the employment status of individual  $i$  living in employment zone  $z(i)$  and neighborhood  $n(i)$  at time  $t$ . It is a dummy variable equal to 1 if the individual is employed, and 0 otherwise (unemployed or inactive).<sup>9</sup>  $DIV_{z(i)t}$  and  $DIV_{n(i)t}$  refer to diversity at the employment zone and neighborhood level at time  $t$ , respectively, the various measures of diversity (nationality-based, birth country-based, and parents' origins-based) being included separately, in different regressions.  $W_{z(i)t}^z$  and  $W_{n(i)t}^n$  are vectors of (average) socioeconomic and demographic characteristics at the employment zone and neighborhood level, respectively.  $X_{it}$  is a set of individual control variables. We also include *département* fixed effects  $\varphi_{d(i)}$ <sup>10</sup> along with time fixed effects  $\varphi_t$  (quarter dummies). Finally,  $\varepsilon_{it}$  is an error term. The two coefficients of interest are  $\beta_1$  and  $\beta_2$ .

9. An alternative dummy variable considered takes value one if the individual is employed and zero if she is unemployed, setting aside inactive individuals. Using this alternative dependent variable does not significantly alter the estimated coefficients.

10. There are 96 *départements* in metropolitan France, corresponding to administrative units—comparable to counties in the United States—usually in charge of managing a number of social and welfare programs, primary and secondary schools, buildings and technical staff, local roads, rural buses, and other large public infrastructure. Some employment zones stretch over several *départements*, but all *départements* contain several employment zones. By contrast, the very local geographical units that we use in the analysis of local diversity, namely local neighborhoods and sectors, are always strictly included in a given *département* (and all *départements* contain several sectors and local neighborhoods).

The inclusion of average socioeconomic and demographic characteristics at both employment ( $W_{z(i)t}^z$ ) and neighborhood ( $W_{n(i)t}^n$ ) levels is meant to account for the fact that diversity could otherwise capture other observable neighborhood characteristics. More precisely, these vectors include a set of dummy variables accounting for the general educational level of the local neighborhood and the employment zone population: share of dropouts, share of individuals having a lower education degree, and share of higher education graduates—with the intermediate high school (*baccalauréat*) graduates as the reference group. We also control for the general social structure of the population (an indirect way of accounting for social capital) by including the share of single-parent households in the local neighborhood and in the employment zone in  $W_{n(i)t}^n$  and  $W_{z(i)t}^z$ , respectively. Finally, these vectors include the shares of individuals with North African (Maghreb) origins, the largest immigrant group in France. It is also important to control for the shares of individuals from Maghreb because this population is more likely subject to employment discrimination, which could affect their employment prospects as well as their neighbors'.<sup>11</sup>

The set of individual controls  $X_{it}$  comprises standard sociodemographic variables: a quadratic function of age, gender, origin, education, and socioeconomic category. The origin variable corresponds to one of the three six-categories measures of origin described in Section III. Specifically, we alternatively include nationality, birth country, and parents' origins indicators when diversity is measured based on nationality, birth country, and parents' origins, respectively. The education variable describes the highest degree obtained by the individual, aggregated into the following four categories: dropouts, lower education, high school, and higher education.<sup>12</sup> We finally control for six possible socioeconomic categories: craftsman, shopkeeper, executive or other high position, intermediate occupation, employee, and factory worker.

To get a first insight on the effects of diversity on employment, we start by estimating a linear probability model.<sup>13</sup> The results presented in Table 4 thus derive from simple OLS estimates, with robust standard errors clustered at the employment zone.<sup>14</sup> For simplicity, Table 4 only reports the estimates for diversity in alternative specifications.<sup>15</sup> The three panels correspond to the alternative measures of diversity (nationality, country of birth, and parents' origins) estimated in separate regressions. In Column 1, we only include time and geographic fixed effects. We progressively add more controls in the subsequent specifications: individual characteristics in Column 2, local average of socioeconomic characteristics in Column 3, and share of North African individuals in the neighborhood and the employment zone in Column 4. Remarkably, the estimates of

11. Given the negligible influence of variables, such as average age and share of males, we decided to report results controlling only for average human capital, family structure, and share of Maghrebians (which are also the most economically relevant characteristics) in the paper.

12. Lower education includes junior high school (BEPC), early vocational training (CAP, BEP); high school includes general and technical high school degrees (Baccalauréat général, technologique ou professionnel); higher education includes Undergraduate (two years after the Baccalauréat), Bachelor's (three years after the Baccalauréat), Graduate (four years after the Baccalauréat), and higher degree (Masters and Ph.D.).

13. Logistic regressions lead to qualitatively similar results, but OLS estimates are displayed for simplicity.

14. The level of clustering (employment zone) corresponds to the highest level of variation of the diversity variable as well as other controls. Employment zone level variables are, by definition, perfectly correlated within employment zone, which magnifies the downward bias of traditional standard errors in case of an even mild within-employment zone correlation in the error term. (Cameron and Miller 2015).

15. The full set of estimates is displayed in Appendix Table A.1 and A.2.

**Table 4**  
*Diversity and Employment: Summary of OLS Regressions*

	Dependent Variable: Employment Status			
	Time and Dep FE (1)	(1)+ Indiv. Char (2)	(2)+ Environment (3)	(3)+ Share North African (4)
<b>Diversity by nationality</b>				
Local neighborhood	-0.260*** (0.026)	-0.164*** (0.021)	-0.114*** (0.017)	-0.092*** (0.014)
Employment zone	0.091*** (0.034)	0.060** (0.029)	0.066** (0.027)	0.095*** (0.034)
<b>Diversity by birth country</b>				
Local neighborhood	-0.225*** (0.014)	-0.130*** (0.014)	-0.090*** (0.011)	-0.072*** (0.009)
Employment zone	0.088*** (0.023)	0.043** (0.018)	0.040** (0.017)	0.058** (0.024)
<b>Diversity by parents' origins</b>				
Local neighborhood	-0.179*** (0.009)	-0.091*** (0.009)	-0.061*** (0.008)	-0.047*** (0.008)
Employment zone	0.043** (0.022)	0.021 (0.015)	0.020 (0.014)	0.025 (0.017)
Individual controls	No	Yes	Yes	Yes
Socioeconomic environment	No	No	Yes	Yes

Notes:  $N=396, 823$ . The sample contains the first and last observations of each sampled individual during years 2007–2013. The sample is described in more detail in Section III. All specifications include quarter and département fixed effects as well as individual and socioeconomic controls. Individual controls include a quadratic function of age and indicators for origin, gender, education, and socioeconomic status. More details on these variables can be found in Tables 1 and 2. Socioeconomic environment controls include the share of dropouts, lower education and higher education graduates, and the share of single-parent household computed both at the employment zone and at the local neighborhood levels. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Robust standard errors in parentheses are clustered at the employment zone level.

local neighborhood diversity are significantly negative for all types of diversity and across all specifications, revealing that local diversity is negatively associated with the probability of being employed. The coefficients are attenuated by the inclusion of all the relevant controls, consistent with the notion that they are correlated to diversity. In particular, controlling for the share of North African individuals significantly reduces the estimates. This does not come as a surprise given the high level of correlation between this variable and the diversity measure (around 60 percent). The estimates for

employment zone diversity all have a significantly positive sign across all specifications, except for diversity based on parents' origins. Interestingly, the correlation between diversity and employment is always the largest (in absolute terms) with the nationality measure and the weakest with the parents' origins measure, suggesting a preeminent role of cultural over ethnic diversity.

There are, however, several reasons to suspect that diversity might actually be endogenous. First, individuals who have a taste for diversity might both self-select into highly diverse areas and be more able or willing to communicate with their neighbors. In this case, people living in more diverse areas should be more likely to communicate with each other, and the negative effect found previously would be overestimated (less negative than the true effect). Reverse causality could also be a problem if immigrants are attracted by more economically dynamic places, where jobs are more abundant. The issue of endogeneity related to the nonrandom location of individuals is addressed in the next two subsections, where we propose alternative identification strategies for the different levels of diversity.

### ***B. Employment Zone Diversity: An Instrumental Variable Approach***

We start by dealing with the issue of endogeneity at the employment zone level. To this aim, we rely on standard instrumental variables estimations. A plausible instrument should be correlated with employment zone level of diversity (for example, employment zone population composition), but uncorrelated to labor market outcomes. We propose two different instruments that we think satisfy these assumptions.

The first instrument relies on the "shift-share" methodology initiated by Card (2001) and more recently used by Saiz (2007) and Ottaviano and Peri (2006) in a setting close to that of this study. It builds on the idea that new migrants to a country tend to settle where former migrants from the same origin previously settled, that is, ethnic enclaves (Munshi 2003; Winters et al. 2001). Using past settlements of immigrants from various countries across French employment zones, it is possible to construct a predicted measure of current diversity in each employment zone. More precisely, we use the 1968 population census data to compute the distribution of each origin group across French employment zones. Because employment zones did not exist in 1968, and because their frontiers evolved over time, we apply the 2010 employment zones borders to the 1968 population. The origin groups considered are limited by the information contained in the 1968 census. In particular, no information about parents is available. We alternatively rely on nationalities and birth countries, grouped into the six categories defined previously. Then, for each origin group, we apply the 1968 distribution of groups across employment zones to the current (2007–2013) total population in France. Doing so, we compute the expected number of individuals from each origin in each employment zone, based solely on the ethnic enclaves pull factor. From this, we can deduce the predicted composition of each employment zone population. Once the predicted shares of each group are computed, we can eventually construct the predicted level of diversity in each employment zone over the 2007–2013 period. By construction, the predicted measure of diversity depends only on 1968 population settlements and not on any employment zone-specific shock (for example, productivity shock) and can thus be used to instrument actual diversity. A more formal description how the predicted level of diversity is computed can be found in Appendix 1.

Although this type of instrument is highly relied upon in the immigration literature, a potential concern may arise if the 1968 distribution of immigrants across zones is related to contemporaneous economic conditions. Immigrants' location in 1968 was actually to a large extent related to regional industrial composition: immigrants were hired in very specific industries that required massive labor force inflows during the post-war reconstruction period (1950s) and the industrialization process (1960s onwards). Immigrants were thus directed towards construction and industry jobs, such that areas characterized by high concentration of jobs in these sectors attracted the bulk of immigrants. To address this issue, we control for the 1968 sectoral composition of the local economy (share of local employment in construction, heavy manufacturing, and light manufacturing) and for the 1968 share of low-skilled workers in the employment zones as potential confounding factors.

The alternative instrument is more innovative and builds on Algan et al. (2016). In this paper, the authors provide evidence that the allocation of households across public housing units within French départements does not take their origins or their preference for diversity into account, so that diversity can be considered as exogenous in the public housing sector. Not only do legal rules prohibit housing allocation based on ethnic backgrounds, but the characteristics of the public housing sector, which is very tight and highly regulated, also make it very complicated to bypass the law in practice. In addition to presenting these general arguments, the authors conduct a variety of formal statistical tests to verify the absence of self-sorting on ethnic characteristics. In particular, they show that the observed spatial distribution of residents across public housing blocks is not statistically different from a random distribution. The level of diversity within the public housing sector of a given area is necessarily correlated to the global level of diversity of this area. Indeed, because the public housing population is part of the total population, fractionalization based only on this subpopulation is mechanically proportional to fractionalization based on the population as a whole. Also, it is reasonable to think that individuals living in the private housing market, who are therefore less constrained upon their location choice, are influenced by the composition of the public housing population. Otherwise stated, people from a given group might be attracted by an area where some of their co-ethnics were located through the public housing allocation process, so that diversity in the area is likely to reflect diversity in the area's public housing sector. In sum, public housing diversity can be considered as exogenous, and it is correlated to general diversity both mechanically and through a magnet effect on immigrants living in the private sector. It can therefore reasonably be used to instrument the general level of diversity.

The effects of diversity on employment are estimated using two-stage least square procedures, where the two instrumental variables described above are alternatively employed. Aside from the inclusion of the instrument, the specification corresponds to our baseline specification.<sup>16</sup> The results are summarized in Table 5, where we only report the coefficients and statistics of interest. For the sake of comparability, we report naive OLS estimates in the first column. Column 2 displays the results obtained using the predicted measure of employment zone diversity as an instrument for employment zone diversity. More precisely, in the first panel of the table, which deals with diversity

---

16. Note that all specifications include quarter fixed effects (2007Q1, 2007Q2, ..., 2013Q4) so that all identifying variation in our IV regressions comes from cross-sectional variation in exposure to the instruments.

**Table 5**  
*Diversity and Employment: IV Regressions*

	Dependent Variable: Employment Status			
	OLS (1)	IV: Expected Diversity (2)	IV: Expected Diversity <sup>a</sup> (3)	IV: Public Housing Diversity (4)
<b>Diversity by nationality</b>				
Employment zone	0.095*** (0.034)	-0.247 (0.214)	-0.210 (0.232)	0.154 (0.121)
Local neighborhood	-0.092*** (0.014)	-0.052* (0.029)	-0.056* (0.030)	-0.101*** (0.018)
First stage				
<i>F</i> -statistic		12.44	10.75	54.42
Coefficient		0.200*** (0.057)	0.208*** (0.064)	0.076*** (0.010)
<b>Diversity by birth country</b>				
Employment zone	0.058** (0.024)	-0.101 (0.111)	-0.050 (0.115)	0.038 (0.071)
Local neighborhood	-0.072*** (0.009)	-0.053*** (0.016)	-0.058*** (0.016)	-0.074*** (0.011)
First stage				
<i>F</i> -statistic		23.42	22.49	87.35
Coefficient		0.440*** (0.091)	0.491*** (0.104)	0.108*** (0.012)
<b>Diversity by parents' origins: Expected diversity based on nationality</b>				
Employment zone	0.025 (0.017)	-0.079 (0.085)	-0.070 (0.096)	0.036 (0.056)
Local neighborhood	-0.047*** (0.008)	-0.034** (0.014)	-0.034** (0.015)	-0.048*** (0.010)
First stage				
<i>F</i> -statistic		32.51	34.58	109.2
Coefficient		0.406*** (0.071)	0.416*** (0.071)	0.126*** (0.012)
<b>Diversity by parents' origins: Expected diversity based on birth country</b>				
Employment zone	-0.026 (0.064)		0.002 (0.071)	
Local neighborhood	-0.041*** (0.011)		-0.043*** (0.012)	

(continued)

**Table 5** (continued)

	Dependent Variable: Employment Status			
	OLS (1)	IV: Expected Diversity (2)	IV: Expected Diversity <sup>a</sup> (3)	IV: Public Housing Diversity (4)
First stage				
<i>F</i> -statistic	65.07		76.39	
Coefficient	0.672*** (0.083)		0.727*** (0.083)	

Notes:  $N=396, 823$ . The sample contains the first and last observations of each sampled individual during years 2007–2013. The sample is described in more detail in Section III. All specifications include quarter and département fixed effects as well as individual and socioeconomic controls. Individual controls include a quadratic function of age and indicators for origin, gender, education, and socioeconomic status. More details on these variables can be found in Tables 1 and 2. Socioeconomic environment controls include the share of dropouts, lower education and higher education graduates, the share of single-parent household, and the share of individuals with North African origins, computed both at the employment zone and at the local neighborhood levels.  $*p < 0.10$ ,  $**p < 0.05$ ,  $***p < 0.01$ . Robust standard errors in parentheses are clustered at the employment zone level.

<sup>a</sup>This column includes employment zones shares of construction and industry workers as well as of individuals with no degree in 1968.

based on nationality, the predicted measure of diversity is also based on nationality. In the second panel of the table dealing with birth country–based diversity, we use predicted diversity based on birth country instead. Finally, in the third panel, given that we do not have any information about parents' origins in the 1968 census, we instrument parents' origins–based diversity with the predicted level of diversity based on nationality and on birth country, alternatively. Information regarding the first stage of the instrumental estimation is reported at the bottom of each panel. It includes the first-stage coefficient (and its standard errors), as well as the associated *F*-statistic. We see that in all cases the excluded instrument displays a strong statistical association with the endogenous variable.<sup>17</sup> Column 2 shows that, for all types of diversity, using the shift-share instrument results in (i) nonsignificant estimates and (ii) a downward adjustment of the effect of diversity on employment. These results are robust to the inclusions of three potential confounding factors accounting for economic conditions that likely explained immigrants' localization in 1968, as reported in Column 3. The results obtained using the second instrument—public housing diversity—reported in Column 4 are also all nonsignificant, although the point estimates are much closer to the OLS ones.

Accounting for the potential endogeneity of location across local labor markets, we do not find robust evidence of a causal impact of employment zone diversity on individuals' employment status. This could lend credence to the suspicion that the naive

17. The *F*-statistics are larger than the rule-of-thumb value of 10 indicated by the literature on weak instruments (Staiger and Stock 1997).

**Table 6**  
*Diversity and Employment: Public Housing Sample (OLS)*

	Dependent Variable: Employment Status	
	All Controls (1)	(1)+ Share North African (2)
Diversity by nationality		
Local neighborhood	-0.153*** (0.019)	-0.159*** (0.020)
Employment zone	0.024 (0.052)	0.017 (0.062)
Diversity by birth country		
Local neighborhood	-0.133*** (0.015)	-0.129*** (0.019)
Employment zone	0.026 (0.040)	0.037 (0.052)
Diversity by parents' origins		
Local neighborhood	-0.100*** (0.016)	-0.091*** (0.018)
Employment zone	0.057 (0.037)	0.062 (0.046)
Observations	52608	52608

Notes:  $N=52,608$ . The sample contains the first and last observations of each sampled individual living in a public housing unit during the years 2007–2013. The sample is described in more detail in Section III. All specifications include quarter and département fixed effects as well as individual and socioeconomic controls. Individual controls include a quadratic function of age and indicators for origin, gender, education, and socioeconomic status. More details on these variables can be found in Tables 1 and 2. Socioeconomic environment controls include the share of dropouts, lower education and higher education graduates, and the share of single-parent household computed both at the employment zone and at the local neighborhood levels. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Robust standard errors in parentheses are clustered at the employment zone level.

estimates reported in the previous section were biased upward, due to a selection of immigrants into more economically dynamic areas. However, the imprecision of the 2SLS estimates precludes any firm conclusion in that respect. Yet, we now turn to the analysis of a subsample, which allows us to find some additional evidence in support of a null effect of aggregate diversity on employment.<sup>18</sup>

18. A comment on local neighborhood diversity is in order here. The corresponding estimates reported in Table 5 are most likely biased due to endogenous selection into local neighborhoods. Yet, we think that running a

Building on Algan et al. (2016), we use the fact that within a given département, people have very little choices in terms of public housing location, thus making the diversity within public housing plausibly exogenous, conditional on département fixed effects.<sup>19</sup> In this context, simple OLS estimates of the effect of diversity on employment should be unbiased. Accordingly, we regress employment on local neighborhood and employment zone diversity, including the full set of controls, restricting our sample to the specific subpopulation of public housing tenants. The results, reported in Table 6 are qualitatively remarkably in line with those obtained on the full sample: We find a significantly negative effect of local neighborhood diversity and no effect of employment zone diversity on employment. An exception is for diversity based on parents' origins, which has a positive effect, though limited and only weakly significant. As expected, the magnitude of the local neighborhood diversity effect is larger than in the general case, given that public housing neighborhoods are generally characterized by larger immigration, lower education, and more generally, by a poorer economic situation. Although these estimates reflect the specificities of a particular subpopulation, they still comfort us on the general direction of the different effects.

### ***C. Neighborhood Diversity: A Local Approach***

We now specifically deal with the endogeneity of local neighborhoods diversity. The approach adopted here builds on the very local nature of the data. It follows Bayer et al. (2008), who study the role of neighbors on work location. The idea is that although households are able to select a given area in which they want to live, they are, however, unable to select a precise neighborhood within this given area. This assumption means that even if households are able to choose a given residential area, there will not be any correlation in unobserved factors affecting employment probability among individuals living in the same neighborhood *within the larger selected area*.<sup>20</sup>

Here, we assume that the households select into *sectors*, the areas encompassing six contiguous local neighborhoods described in Section III, that are containing on average 120 households. Several standard arguments support the assumption that even if households were able to select a precise sector, they would not be able to choose among the various local neighborhoods within this selected sector. First, because the housing market is very tight, it is reasonable to think that an individual targeting a given sector is very unlikely to have a choice over the precise neighborhood where she will end up in this area. This would indeed require that at least one housing unit satisfying the other decision criteria of the individual (for example, size) be vacant in each local neighborhood within the target sector at the time when the individual is searching for a new

---

analysis strictly focused on employment zone diversity (that is, excluding local neighborhood diversity) would be inappropriate as it would present a risk of omitted variable bias. As we explain in Section II, diversity could influence employment at the local (neighborhood) or aggregate (employment zone) levels through different mechanisms. Excluding local diversity from the regression would imply that the coefficient for diversity at the aggregate level also absorb the effect of local diversity. The next subsection specifically deals with endogeneity of local neighborhood diversity. Yet, we now propose a side-specification in which diversity is exogenous at both aggregate and local levels, although in a very specific context.

19. The public housing diversity instrument is also based on this intuition.

20. Schmutte (2015) follows a similar approach to study the propensity of neighbors to share information about high employers. We carry out a sorting test on observables in the spirit of that paper.

place. A second consideration is that it may be difficult for prospecting individuals to identify neighborhood-by-neighborhood variation in neighbors and contextual characteristics, prior to moving into the neighborhood. To put it differently, although the individual may have a realistic ex ante view on the characteristics of the targeted sector, it is less likely that she is actually able to identify differences in these characteristics across the various neighborhoods of the sector. This is particularly reasonable in our case, as the local neighborhoods considered consist of about 20 households. Finally, the local neighborhoods studied here do not correspond to any administrative or official frontiers (nor do sectors). People do not know where the borders are, and more generally, do not even know what a local neighborhood (*aire* in the LFS) is because it is only used as the sampling unit of the LFS.

For those reasons, it is close to impossible that French households purposely decide to live in a given local neighborhood rather than the next one in a given sector. All these arguments support the validity of the assumption that there should be no correlation in unobserved factors affecting employment among neighbors living in the same neighborhood within the same sector. As a consequence, once we control for the characteristics of the sector selected by the individual, the remaining spatial variance of diversity across neighborhoods within the sector is supposed to be exogenous. This is done through the inclusion of sector fixed effects: In this case the comparison group for individuals living in a given local neighborhood is made of all individuals living in the same sector but in a different local neighborhood.

A potential limit of this approach can come from the fact that, by construction, the comparison group includes individuals who are not surveyed at the same time, although they live in a very limited geographic zone (in cross section). In other words, when we include sector fixed effects, we use as a “control group” individuals who live in the same place (same sector but different local neighborhood), but at different points in time. Note that all aggregate shocks that have a uniform effect across periods are absorbed by the inclusion of time (quarter) fixed effects. This ensures that our identifying variation does not rely directly on variation in employment probability across periods. Yet, it does not solve the problem of time-varying shocks that could affect different communities differently. To bypass this issue, we propose a more constraining alternative strategy whereby we include fixed effects for an even more restrictive zone, which we call *subsector*. We define a subsector as the set of these two local neighborhoods in a given sector which are surveyed exactly one after the other: Individuals living in the same subsector are those surveyed in the last (sixth) wave of a local neighborhood and those surveyed in the first wave of the adjacent local neighborhood.<sup>21</sup> With this strategy, the reference group is made of individuals who are not only observed living in two contiguous local neighborhoods of the same sector, but who are also observed living there at the same time (at most one quarter apart). It is precisely to allow for the inclusion of these subsectors that we needed to restrict our sample to the the first and last (sixth) wave of each local neighborhood, as mentioned in Section III.<sup>22</sup>

21. Consider, for instance, local neighborhood  $z$ ,  $LN_z$ , which is surveyed six quarters in a row:  $LN_z^1, LN_z^2, \dots, LN_z^6$ . Afterwards, the adjacent local neighborhood,  $LN_{z+1}$  is included for the next six waves of the survey:  $LN_{z+1}^1, LN_{z+1}^2, \dots, LN_{z+1}^6$ . The last wave of a given local neighborhood,  $LN_z^6$ , thus takes place exactly one quarter before the first wave of the subsequent local neighborhood,  $LN_{z+1}^1$ .

22. Obviously, the first wave of survey of the first local neighborhood of a given sector cannot be matched with a previous one, nor can the last wave of the last local neighborhood be matched with a next one. Therefore, we

In addition to the general arguments supporting the strategy presented above, we now provide a more formal sorting test to verify that there is very little remaining sorting on the basis of observable characteristics once we control for selection into higher level areas (sectors and subsectors). The general idea of the test is to assess whether individuals systematically tend to locate in neighborhoods where residents are similar to themselves regarding a number of observable characteristics (such as educational attainment or origin for instance). If, as we argue, households select their residence at the level of the (sub)sector (at worst), but do not pick a specific location (here, a local neighborhood) within this (sub)sector, then we should not observe much systematic correlation between local neighbors' observable characteristics within a local neighborhood, once sorting across (sub)sectors is accounted for, through the inclusion of (sub)sector fixed effects. The statistical test that we propose here, directly inspired by Bayer et al. (2008), thus consists of checking that once one conditions on fixed effects for the supposed area of selection—(sub)sector in our case—the social makeup of the local neighborhood is not strongly predictive of individual background.

As a first step of this test, we verify that there is indeed spatial sorting at larger geographic levels. To this aim, we randomly pick one individual per local neighborhood and regress any of her observable characteristics ( $D_{i,n}$ ) on the local neighborhood average of this characteristic, excluding the individual and members of her household ( $\bar{D}_{-in}$ ).<sup>23,24</sup> The  $R$ -squared from these regressions for different observable characteristics are reported in Column 1 of Table 7. As expected, they are rather large, suggesting a substantial degree of spatial sorting.

In a second step, we want to test the assumption that, while people might be able to sort into a given (sub)sector, they cannot select a specific local neighborhood *within* this chosen (sub)sector. To do so, we first regress the individual characteristic ( $D_{i,n}$ ) onto a set of fixed effects for alternative geographic areas, from the larger to the smaller: départements, sectors, and subsectors. We denote  $\tilde{D}_{i,n}$  the residuals from these regressions. We then run the same regressions using instead the corresponding neighborhood average ( $\bar{D}_{-i,n}$ ) as the dependent variable, from which we obtain residuals noted  $\tilde{D}_{-i,n}$ . Both residuals ( $\tilde{D}_{i,n}$  and  $\tilde{D}_{-i,n}$ ) imply measure deviation from the area average. If, as stated above, households are unable to select which particular local neighborhood they settle in *within* a given area, there should be no association between the residual variables: ( $\tilde{D}_{-i,n}$ ) should not be a good predictor of  $\tilde{D}_{i,n}$ . The  $R$ -squared from the regressions of  $\tilde{D}_{i,n}$  on  $\tilde{D}_{-i,n}$  are reported in Columns 2–4 of Table 7. The “Fixed Effects” indicated at the top of each column refers to the set of geographical fixed effects onto which we regress alternatively  $D_{i,n}$  and  $\bar{D}_{-i,n}$ . The results presented in Column 2 are obtained using département fixed effects. Départements are rather large administrative areas (96 départements in mainland France) across which and within which sorting is very likely to occur. Therefore, it is not surprising that the  $R$ -squared remain very similar to those in Column 1. By contrast, including sector or subsector fixed effects (Columns 3 and 4,

have two observations for people living in  $LN_2$  to  $LN_5$ , and only one for those living in  $LN_1$  ( $LN_1^6$ ) and  $LN_6$  ( $LN_6^1$ ).

23. The draw of a single individual in the neighborhood is intended to avoid the systematic negative bias related to the fact that neighborhood averages of covariates exclude the reference individual, as explained in Bayer et al. (2008).

24. In order to make sure our results do not depend on a specific drawing, we repeat the drawing 100 times and report average  $R$ -squared for the various steps of this test.

**Table 7**  
*Sorting Test: R-squared (in Percent)*

Variable	Unconditional (1)	Département Fixed Effects (2)	Sector Fixed Effects (3)	Subsector Fixed Effects (4)
<b>Nationality</b>				
French	1.83	1.01	0.19	0.26
South European	0.15	0.11	0.03	0.04
Other European	0.10	0.07	0.02	0.04
North African	0.66	0.53	0.07	0.04
Other African	0.28	0.20	0.13	0.15
Other nationality	0.15	0.09	0.03	0.02
<b>Education</b>				
Master, Ph.D.	4.91	2.15	0.11	0.10
Graduate	0.18	0.11	0.04	0.03
Undergraduate	0.07	0.05	0.01	0.03
Lower undergraduate	0.32	0.28	0.07	0.08
Technical Baccalaureate	0.16	0.10	0.03	0.03
Lower degree	1.84	0.90	0.19	0.17
Dropout	2.71	2.07	0.22	0.23
<b>Other</b>				
Age	2.72	2.40	0.70	0.71

Notes: In each neighborhood, we randomly draw an individual and compute the average characteristics among her neighbors (excluding members of her own household). For each observed characteristic, we regress the individual's observation on neighbors' average observation (including quarter fixed effects) and look at the within-group *R*-squared of the regression. We repeat this process 100 times and compute the average *R*-squared, which are reported in the table, in percent. Column 1 displays the average *R*-squared from the unconditional regressions, while Columns 2, 3, and 4 display the mean *R*-squared from regressions including département, sector, and subsector fixed effects, respectively. The sample includes working age, nonstudent individuals from 2007 to 2013, in the first and sixth waves of neighborhoods' sampling.

respectively) leads to a substantial decrease in *R*-squared. For instance, the average *R*-squared of 4.9 percent reported in Column 1 for the "Master, Ph.D." variable clearly reveals residential sorting among high-skilled individuals, who tend to live in areas with a high share of high-skilled neighbors. The *R*-squared remains high when considering within-département sorting, at about 2.15 percent in Column 2. However, conditioning on sector or subsector fixed effects drives the *R*-squared down to 0.1 percent (Columns 3 and 4). This supports the assumption that the amount of within-(sub)sector sorting based on higher education is negligible. The same generally holds for relevant observable characteristics such as other levels of education, age, and the various nationalities.

Table 8 reports the estimates of local neighborhood diversity correcting for sorting through the inclusion of (sub)sector fixed effects instead of département fixed effects. The specification used is the baseline specification including the full set of control, and

**Table 8**  
*Diversity and Employment at the Local Neighborhood Level: Local Fixed Effects Method*

	Dependent Variable: Employment Status		
	(1)	(2)	(3)
Diversity by nationality—local neighborhood	−0.080*** (0.012)	−0.058*** (0.009)	−0.059*** (0.011)
Diversity by birth country—local neighborhood	−0.064*** (0.009)	−0.046*** (0.008)	−0.048*** (0.009)
Diversity by parents' origins—local neighborhood	−0.042*** (0.008)	−0.029*** (0.008)	−0.029*** (0.009)
Département FE	Yes	No	No
Sector FE	No	Yes	No
Sub Sector FE	No	No	Yes

Notes:  $N=396, 823$ . The sample contains the first and last observations of each sampled individual during years 2007–2013. The sample is described in more detail in Section III. All specifications include quarter fixed effects as well as individual and socioeconomic controls. Individual controls include a quadratic function of age and indicators for origin, gender, education, and socioeconomic status. More details on these variables can be found in Tables 1 and 2. Socioeconomic environment controls include the share of dropouts, lower education and higher education graduates, the share of single-parent household, and the share of individuals with North African origins, computed both at the employment zone and at the local neighborhood levels.  $*p < 0.10$ ,  $**p < 0.05$ ,  $***p < 0.01$ . Robust standard errors in parentheses are clustered at the employment zone level.

including alternatively sector fixed effects (Column 2) and subsector fixed effects (Column 3). Note that this strategy forces the exclusion of employment zone diversity, since all the variation at this aggregate level is captured by the inclusion of (sub)sector fixed effects. However, excluding this variable does not put the specification at risk of any omitted bias. In addition, evidence presented in the previous section suggests that aggregate diversity does not have a systematic employment effect. For this reason and for the sake of comparability, Column 1 reports the naive estimates with département fixed effects excluding employment zone diversity. The estimates for local neighborhood diversity reported in Columns 2 and 3 are very similar, suggesting similar patterns of selection into sectors and subsectors. This is not surprising given the low residential mobility, especially over short periods of time. The coefficient for all types of diversity are still significantly negative, although smaller (in magnitude) than in the naive regressions. This suggests that individuals sort across neighborhoods in such a way that more economically deprived individuals live in more diverse neighborhoods. A substantial result emerges from comparing the coefficients for the various measures of diversity: The estimated effect of nationality-based diversity (−0.058) is larger (in absolute terms) than that of birth country-based diversity (−0.048), which is also larger than parents' origins-based diversity (−0.029). When we measure the effect in terms of standard deviation of the corresponding diversity measure, the differences

are less pronounced because diversity by parents' origins is much more dispersed than diversity based on nationality, but this ranking remains unchanged: A one standard deviation increase in neighborhood diversity based on nationality leads to a 0.83 percent decrease in employment probability, representing 1.25 percent of the mean employment probability. The corresponding figure for local diversity based on parents' origins is 0.7 percent.

These results suggest that living in a context where people have different nationalities matters more for employment than living in a context where people were born in different countries, and even more than living close to people whose parents are from different origins. As discussed in Section III, parents' origins more likely reflect ethnicity than nationality, the latter being rather associated to common values and language. An interpretation of these results is therefore that neighborhood diversity lowers the probability of employment because of cultural differences, most likely including language differences, rather than ethnic differences. This reinforces the intuition that diversity affects job finding by limiting job information transmission between neighbors, which we try to test in the next section.

## V. Interpretation of the Results

So far, we have shown that the positive relationship between employment zone diversity and individuals' employment probability is no longer significant once we account for endogeneity. This observation, combined with our analysis on the specific sample of public housing tenant, suggests the absence of a causal relationship between diversity and employment at aggregate levels. By contrast, we have also established that living in a diverse neighborhood actually implies a lower employment probability, an effect that remains once we correct for residential sorting. This section explores the potential mechanisms by which diversity reduces individuals' employment prospects at the very local level. As discussed in Section II, one of the possible channels underlying the relationship between neighborhood diversity and employment is that of communication and job information transmission between agents. Specifically, if neighbors from different origins do not communicate—for example, because they do not speak the same language or do not share common social norms—then information about job opportunities or about how to register at an employment agency does not circulate across groups.

One of the results obtained in the previous section is that the negative effect of local neighborhood diversity on employment is particularly salient when we consider diversity by nationality rather than diversity based on parental origins or birth country. Given our postulate that a common nationality is likely to reflect a common cultural identity (for example, shared language and set of social norms),<sup>25</sup> this result is in line with the idea that the effect of diversity at the local level goes through the channel of less communication among more diverse communities. This section aims to test this assumption more formally.

---

25. We insist that this common cultural belonging conferred by a common nationality may come on top of another cultural background or of a different ethnicity inherited from the parents or the country of birth. We do not overlook this possibility, but simply focus on the shared characteristics going with a shared nationality.

A first test of this potential mechanism consists of looking at how employment status is affected by the presence of people from the same origin group. Presumably, if the negative effect of neighborhood diversity is due to limited information transmission across groups, then being surrounded by neighbors with whom one is more likely to communicate should conversely be related to better employment prospects. For each individual we compute the share of her local neighbors belonging to the same origin group—nationality, birth country or parents' origins—and regress her employment status on this variable, including the full set of controls and (sub)sector fixed effects. The estimated coefficients on the share of individuals from one's own group are reported in Table 9. A first glance at Columns 1 and 2, where diversity is excluded from the regression, shows that a larger share of neighbors from one's own group is associated with a higher employment probability. This is especially true when the individual's origin group is measured by her nationality. This tends to support the notion that the more one is able to communicate with one's neighbors (for example, when they have a common origin, and especially when they share the same nationality), her employment prospects are improved. Because a higher level of local diversity is on average associated with a lower share of neighbors from one's own group, these results could explain why diversity reduces employment. Yet, when we run the baseline regressions of employment on diversity, controlling for the shares of neighbors from the same group (Columns 3 and 4), we observe that the negative effect of diversity is not annihilated. This shows that diversity has an adverse effect on labor market outcomes above and beyond its impact through its negative relationship with the share of same-group neighbors. In addition, we see that the estimated coefficient on own-group share vanishes once diversity is accounted for, except in the case of nationality. This suggests that in the case of nationality only, part of the negative effect of diversity is mediated through the own-group share channel. This result reinforces the idea that nationality captures a crucial dimension of origin that is not taken into account by the other definitions and that individuals are most penalized when they do not share a common culture.

The previous test was based on the idea that the negative effect of local neighborhood diversity comes from more difficult communication across culturally heterogeneous groups. The main argument supporting this notion relates to our claim that a common nationality reflects shared cultural norms on top of any other inherited cultural or ethnic belonging. However, one could argue that the estimated effect of diversity by nationality is biased due to the potential endogeneity of the naturalization process. Indeed, if naturalization is the outcome of a highly selective process whereby only the most integrated and motivated foreign-born individuals manage to become French, then unobservable individual characteristics would correlate both with the level of diversity based on nationality and employment outcomes. Existing work however suggests that selection on unobservables into naturalization is not empirically very influential when analyzing the labor market outcomes of foreign-born individuals and that if anything this selection is negative.<sup>26</sup> Yet, we build an alternative classification of individuals' origins that

---

26. Notably Fougère and Safi (2009) show that there is a very small degree of selection on unobservables in the naturalization process and that foreign-born individuals most likely to become French through naturalization tend to have unobservable characteristics that make them less employable (controlling for similar characteristics as those we use in our study). Although this result does not rule out the concern of an endogeneity bias related to selection into naturalization, it suggests that this bias is likely to be very small, and if anything, goes in the opposite direction. Given that the main difference between nationality and birth country is related to

**Table 9**  
*Diversity and Employment: Including the Share of Own Group*

	Dependent Variable: Employment Status			
	(1)	(2)	(3)	(4)
Diversity by nationality				
Share of own group	0.078*** (0.012)	0.074*** (0.012)	0.052*** (0.017)	0.054*** (0.017)
Local neighborhood diversity			-0.033** (0.013)	-0.030** (0.014)
Diversity by birth country				
Share of own group	0.022*** (0.007)	0.018** (0.007)	-0.009 (0.010)	-0.007 (0.010)
Local neighborhood diversity			-0.054*** (0.010)	-0.052*** (0.011)
Diversity by parents' origins				
Share of own group	0.012* (0.006)	0.010 (0.007)	-0.004 (0.007)	-0.003 (0.008)
Local neighborhood diversity			-0.036*** (0.008)	-0.035*** (0.009)
Sector FE	Yes	No	Yes	No
Sub Sector FE	No	Yes	No	Yes

Notes:  $N=396, 823$ . The sample contains the first and last observations of each sampled individual during years 2007–2013. The sample is described in more detail in Section III. All specifications include quarter fixed effects as well as individual and socioeconomic controls. Individual controls include a quadratic function of age and indicators for origin, gender, education and socioeconomic status. More details on these variables can be found in Tables 1 and 2. Socioeconomic environment controls include the share of dropouts, lower education and higher education graduates, the share of single-parent household, and the share of individuals with North African origins, computed both at the employment zone and at the local neighborhood levels. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Robust standard errors in parentheses are clustered at the employment zone level.

reflects their cultural integration without introducing bias related to their propensity to engage successfully in the naturalization process. This new classification allows us to circumvent the potential issues related to the endogeneity of naturalization, and it also permits us to isolate further the role of local cultural diversity in explaining individual labor market outcomes.

The new classification is based on time spent in France by foreign-born individuals, which we use as a proxy for exposure to French culture: The longer an individual has

---

naturalization, this suggests that the “true” difference between the effects of these two types of diversity would actually be larger than the one we measure in our study, rather than smaller.

lived in France, the more likely she is to be impregnated with French culture, including language, social norms, and knowledge of the institutions. For each foreign-born individual, we compute the share of her life spent in France, based on her year of arrival in France.<sup>27</sup> On the basis of this measure, we define an individual's cultural group as her birth country<sup>28</sup> if she has spent less than half of her life in France (even if she has been naturalized), and as France otherwise. Using this measure, 92 percent of our sample belongs to the French cultural group. This group thus includes, among others, foreigners who are not naturalized (whatever the reason) but who have spent a substantial part of their life in France, while it excludes naturalized individuals who have spent less than 50 percent of their life in France. We eventually compute an alternative measure of diversity based on these groups of cultural belonging. Compared to nationality, this measure is less likely to be influenced by integration efforts. Compared to country of birth, it is more stringent because as it classifies foreign born who spent a large share of their life in France as French. Consequently, individuals belonging to the same foreign group of origin are likely to be much closer culturally than in the case of birth country.

We regress employment on this new measure of cultural diversity, at the neighborhood level, including sector fixed effects and controlling for the same set of controls as in the usual specification. The coefficients are reported in Appendix Table A.2. Strikingly, these coefficients are even more negative than the ones obtained for nationality. These results reinforce the notion that cultural diversity is particularly important in explaining employment prospects at the neighborhood level, in line with the idea that the exchange of information about job opportunities may be hindered by a lack of communication across cultural groups.<sup>29</sup>

As an alternative way of testing the impact of local diversity on the quality of neighborhood social capital, we now propose to look at the correlation between local diversity and the nature of neighborhood relationships using the 2002 French Housing Survey. Individuals are asked to qualify the relationships with their neighbors as good, average, bad, or nonexistent. In addition, we know the precise (block level) place of living of the individuals, and we are able to match it with representative block level measures of diversity from the 1999 population census. The results of multinomial logit regressions of the quality of neighborhood relationships on neighborhood diversity are presented Table 10. Each row corresponds to a separate regression: The first row displays the estimated coefficients of diversity based on nationality at birth, those for birth country-based diversity being reported in the second row. Each regression controls for individual characteristics (age, gender, origin, employment status, education, household income), block level unemployment rate, département fixed effects, and a detailed indicator of the social and economic composition of neighborhood in 27 categories.<sup>30</sup> The results reveal that individuals living in more diverse neighborhoods are less likely to report having good relationships with their neighbors. In particular, they are more likely to report

27. For people born in France, we assume that they have spent their whole life in France, so that their exposure is 100 percent. In comparison, we obtain that the average immigrant has spent 56 percent of her life in France.

28. We rely on the five same origin categories as in the rest of the study: southern Europe, rest of Europe, North Africa, rest of Africa and other countries.

29. These results also suggest that the difference between the effect of birth country diversity and nationality diversity was probably underestimated, as discussed Footnote 26.

30. The socioeconomic classification of French neighborhoods into 27 groups is realized by Martin-Houssart and Tabard (2002).

**Table 10**  
*Quality of the Relationships with Neighbors and Diversity*

Variable	Quality of Neighborhood Relationships (Ref: Good)		
	Average	Bad	No Relationship
Diversity by nationality	1.435*** (0.215)	1.707*** (0.551)	0.883*** (0.209)
Diversity by birth country	1.621*** (0.290)	1.958** (0.655)	1.208*** (0.254)

Notes:  $N=28,716$ . Each row reports the coefficients from a separate multinomial logit regression. The dependent variable indicates opinion about the relationships with the neighbors. It takes value 1 if the surveyed individual declares having good relationships with her neighbors (reference category), 2 if the relationships are average, 3 if they are bad, and 4 if there is no relationship at all. The main variable of interest is the level of diversity, computed at the block level, based on nationalities in the first regression and on birth countries in the second one. In each specification, the following controls are included: individual characteristics (age, gender, origin, employment status, education, household income), block level unemployment rate, département fixed effects, and a detailed indicator of the social and economic composition of the neighborhood (27 categories). The data come from the 2002 French Housing Survey and the 1999 population census (INSEE). Standard errors clustered at the neighborhood level are reported in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

having bad relationships than average relationships and than no relationship at all. These simple results tend to support the idea that communication can be hindered in more diverse neighborhoods due to the poor quality of the relationships between neighbors.

Eventually, the French LFS also provides information on job search and job finding methods that enable us to explore the relationship between job search and finding methods and local diversity using the same set of controls as in our baseline specification, and including sector fixed effects. We find no evidence of a relationship between local diversity and the use of informal job search method, defined as relying (exclusively or not) on one's personal network. Yet, individuals' origins matter to some extent in explaining the use of networks to search for a job. People with Mediterranean, Maghrebian, and other African origins (taking second generation into account) are more likely to rely on personal networks than natives. Interestingly, European (other than South European) and African citizens are also more likely to rely exclusively on networks, revealing a low level of integration for those particular groups.<sup>31</sup>

Then, focusing on employed individuals, we find no relationship between the local diversity and the probability of having found a job through one's personal network. However, any employed foreigner is more likely to have found her job using networks than French citizens. This is especially true for individuals of South European and Rest of the World nationalities. The coefficients decrease (or even vanish) as other measures of origins are considered, suggesting that networks are particularly helpful for the least integrated people, that is, those who are of foreign origin but who have not yet been naturalized.<sup>32</sup> Although these results are to be interpreted with caution because they do

31. Results are not reported here but are available upon request.

32. Results are not reported here but are available upon request.

not correct for selection and do not control for the search methods that were actually used, they suggest that friends and family network is an important vector of employment for foreign individuals. Therefore, even if diversity is not directly involved in the use of networks to search and find jobs, it might still be an issue for minorities if they live in diverse areas, isolated from the core of their network.

## VI. Conclusion

The findings of this paper bring new insights to the literature on diversity. First, measuring diversity based on various definitions of origins reveals that diversity in terms of nationalities matters more than diversity in terms of parents' origins. This is a key result, as it means that diversity of origins plays a role through the variety of cultures and languages rather than through ethnic diversity per se. This speaks in favor of the idea that diversity affects employment prospects by altering job information transmission. Second, measuring diversity at different geographic levels reveals that this effect is not independent from the level of observation. Neighborhood diversity reduces employment prospects, while employment zone diversity is neutral, after correcting for endogenous sorting. This implies that the mechanisms through which diversity hinders employment at a local level are counterbalanced at a more aggregate level. In particular, job seekers might be unable to develop efficient networks in their own neighborhood because of diversity, but they might instead rely on a network established in a larger area. More generally, this work calls for a new approach of the literature on diversity, as it shows that (i) the notion of diversity hides various aspects that can influence the outcome considered in different ways and (ii) that the effect of diversity can vary according to the geographical level considered.

Although part of this paper is devoted to testing the hypothesis that the negative impact of local diversity on employment prospects is related to job information transmission, much remains to be done in this direction. In addition, a natural subsequent question is that of the quality of the job found, in terms of tenure or wage for instance. These issues remain open for future research.

## Appendix 1

### Construction of the Predicted Level of Employment Zone Diversity

Denote  $N_{France,1968}^g$  the number of individuals from origin group  $g = 1, \dots, g_{max}$  in France in 1968 and  $N_{EZ_j,1968}^g$  the number of individuals from origin group  $g = 1, \dots, N_g$  in employment zone  $j = 1, \dots, N_j$  in 1968. Then, the share of group  $g$  individuals, living in employment zone  $j$  in 1968 (out of the total number of group  $g$  individuals in France in 1968) can be computed as follows:

$$(3) \quad S_{EZ_j,1968}^g = \frac{N_{EZ_j,1968}^g}{N_{France,1968}^g}$$

with  $\sum_{j=1}^N S_{EZ_j,1968}^g = 1$ , for any group  $g$ .

Then, the expected number of group  $g$  individuals living in employment zone  $j$  in year  $t=2007, \dots, 2010$  is given by:

$$(4) \quad \widehat{N}_{EZj,t}^g = S_{EZj,1968}^g \times N_{France,t}^g$$

From this, we can deduce the expected share of group  $g$  individuals in employment zone  $j$  in year  $t$  (out of the total number of individuals living in employment zone  $j$  in  $t$ , all groups included):

$$(5) \quad \widehat{s}_{EZj,t}^g = \frac{\widehat{N}_{EZj,t}^g}{\sum_{g=1}^{N_g} \widehat{N}_{EZj,t}^g}$$

with  $\sum_{g=1}^{N_g} \widehat{s}_{EZj,t}^g = 1$ , for any employment zone  $j$ .

Finally, the predicted measure of diversity in employment zone  $j$  in  $t$  is obtained as follows:

$$DIV_{EZj,t} = 1 - \sum_{g=1}^{N_g} \left[ \widehat{s}_{EZj,t}^g \right]^2$$

## Appendix 2

### Tables

**Table A.1**

*Diversity (based on nationality) and Employment: Detailed OLS Results*

	Dependent Variable: Employment Status
<b>Diversity</b>	
Local neighborhood (LN)	-0.092*** (0.014)
Employment zone (EZ)	0.095*** (0.034)
<b>Individual characteristics</b>	
Age	0.075*** (0.002)
Age square	-0.001*** (0.000)
Male	0.079*** (0.003)

(continued)

	Dependent Variable: Employment Status
<b>Nationality</b>	
French	0.125*** (0.013)
South Europe	0.191*** (0.011)
Rest of Europe	0.004 (0.013)
Maghreb	-0.016 (0.012)
Rest of Africa	0.005 (0.016)
Rest of the world	(Ref)
<b>Education</b>	
No diploma	-0.090*** (0.004)
Lower education	-0.021*** (0.002)
Higher education	0.024*** (0.002)
High school degree	(Ref)
<b>Occupation</b>	
Craftsman, shopkeeper	0.690*** (0.011)
Executive or other high position	0.671*** (0.008)
Intermediate occupation	0.639*** (0.008)
Employee	0.621*** (0.010)
Factory worker	0.576*** (0.009)
Unemployed worker that has never worked	(Ref)
<b>Socioeconomic environment controls</b>	
Percentage no degree (in LN)	-0.112*** (0.010)
Percentage degree less than high school (in LN)	-0.011 (0.010)

	Dependent Variable: Employment Status
Percentage college degree (in LN)	−0.037*** (0.009)
Percentage single-parent household (in LN)	−0.090*** (0.013)
Percentage North African (in LN)	−0.106 (0.030)
Percentage no degree (in EZ)	−0.037 (0.034)
Percentage degree less than high school (in EZ)	−0.063*** (0.028)
Percentage college degree (in EZ)	−0.027 (0.029)
Percentage single-parent households (in EZ)	−0.084*** (0.032)
Percentage North African (in EZ)	−0.196* (0.112)
Observations	396,813
Quarter FE	Yes
Department FE	Yes

Notes:  $N=396, 823$ . The sample contains the first and last observations of each sampled individual during years 2007–2013. The sample is described in more details in Section III.

**Table A.2**

*The Role of Language and Common Norms: A Test Using a Measure of Exposure to France (Local Fixed Effects Approach)*

	Dependent Variable: Employment Status	
	(1)	(2)
Diversity based on fraction of lifetime spent in France—local neighborhood	−0.076*** (0.010)	−0.074*** (0.012)
Sector FE	Yes	No
Subsector FE	No	Yes

Notes:  $N=396, 823$ . The sample contains the first and last observations of each sampled individual during years 2007–2013. The sample is described in more details in Section III. All specifications include quarter fixed effects as well as individual and socioeconomic controls. *Individual controls* include a quadratic function of age and indicators for origin, gender, education, and socioeconomic status. More details on these variables can be found in Tables 1 and 2. *Socioeconomic environment controls* include the share of dropouts, lower education, and higher education graduates; the share of single-parent household; and the share of individuals with North African origins, computed both at the employment zone and at the local neighborhood levels. \* $p < 0.1$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Robust standard errors in parentheses are clustered at the employment zone level.

## References

- Alesina, Alberto, Reza Baqir, and William Easterly. 1999. "Public Goods and Ethnic Divisions." *The Quarterly Journal of Economics* 114(4):1243–84.
- Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." *Journal of Economic Growth* 8(2):155–94.
- Alesina, Alberto, Johann Harnoss, and Hillel Rapoport. 2016. "Birthplace Diversity and Economic Prosperity." *Journal of Economic Growth* 21(2):101–38.
- Alesina, Alberto, and Eliana La Ferrara. 2000. "Participation in Heterogeneous Communities." *The Quarterly Journal of Economics* 115(3):847–904.
- . 2002. "Who Trusts Others?" *Journal of Public Economics* 85(2):207–34.
- . 2005. "Ethnic Diversity and Economic Performance." *Journal of Economic Literature* 43(3):762–800.
- Algan, Yann, Camille Hémet, and David D. Laitin. 2016. "The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation." *Journal of Political Economy* 124(3):696–733.
- Ananat, Elizabeth, Shihe Fu, and Stephen L. Ross. 2013. "Race-Specific Agglomeration Economies: Social Distance and the Black–White Wage Gap." NBER Working Paper 18933. Cambridge, MA: National Bureau of Economic Research.
- Battu, Harminder, Paul Seaman, and Yves Zenou. 2011. "Job Contact Networks and the Ethnic Minorities." *Labour Economics* 18(1):48–56.
- Bayer, Patrick, Stephen Ross, and Giorgio Topa. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy* 116(6):1150–96.
- Bellini, E., G.I.P. Ottaviano, D. Pinelli, and G. Prarolo. 2013. "Cultural Diversity and Economic Performance: Evidence from European Regions." In *Geography, Institutions and Regional Economic Performance*. Advances in Spatial Science. ed. R. Crescenzi and M. Percoco, 121–41. Berlin: Springer.
- Bisin, Alberto, Eleonora Patacchini, Thierry Verdier, and Yves Zenou. 2011. "Formation and Persistence of Oppositional Identities." *European Economic Review* 55(8):1046–71.
- Borjas, G.J. 2003. "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market." *Quarterly Journal of Economics* 118(4):1335–74.
- Cahuc, P., and S. Carcillo. 2014. "The Detaxation of Overtime Hours: Lessons from the French Experiment." *Journal of Labor Economics* 32(2):361–400.
- Calvo-Armengol, A., and M.O. Jackson. 2004. "The Effects of Social Networks on Employment and Inequality." *American Economic Review* 94(3):426–54.
- Cameron, A.C., and D. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50(2):317–72.
- Card, D. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics* 19(1):22–64.
- Coleman, J.S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94:S95–S120.
- Corcoran, M., L. Datcher, and G. Duncan. 1980. "Information and Influence Networks in Labor Markets." *Five Thousand American Families: Patterns of Economic Progress* 7:1–37.
- Costa, D.L., and M.E. Kahn. 2003. "Civic Engagement and Community Heterogeneity: An Economist's Perspective." *Perspectives on Politics* 1(1):103–111.
- Easterly, W., and R. Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *The Quarterly Journal of Economics* 112(4):1203–1250.
- Fougère, D., and M. Safi. 2009. "Naturalization and Employment of Immigrants in France (1968–1999)." *International Journal of Manpower* 30(1/2):83–96.

- Glennerster, R., E. Miguel, and A.D. Rothenberg. 2013. "Collective Action in Diverse Sierra Leone Communities." *The Economic Journal* 123(568):285–316.
- Granovetter, M.S. 1995. *Getting a Job: A Study of Contacts and Careers*. 2nd edition. Cambridge, MA: Harvard University Press.
- Hellerstein, J.K., D. Neumark, and M. McInerney. 2008. "Spatial Mismatch or Racial Mismatch?" *Journal of Urban Economics* 64(2):464–79.
- Holzer, H.J. 1988. "Search Method Use by Unemployed Youth." *Journal of Labor Economics* 6(1):1–20.
- Hong, L., and S.E. Page. 2001. "Problem Solving by Heterogeneous Agents." *Journal of Economic Theory* 97(1):123–63.
- Ioannides, Y.M., and L. Datcher Loury. 2004. "Job Information Networks, Neighborhood Effects, and Inequality." *Journal of Economic Literature* 42(4):1056–1093.
- Ioannides, Y.M., and G. Topa. 2010. "Neighborhood Effects: Accomplishments and Looking Beyond Them." *Journal of Regional Science* 50(1):343–62.
- Lazear, E.P. 1999. "Globalisation and the Market for Teammates." *Economic Journal* 109(454): C15–40.
- Light, I. 2005. "The Ethnic Economy." In *The Handbook of Economic Sociology*, 2nd edition, ed. Neil J. Smelser and Richard Swedberg, 670–677. Princeton, NJ: Princeton University Press.
- Light, I., and C. Rosenstein. 1995. *Race, Ethnicity, and Entrepreneurship in Urban America*. Berlin: Walter de Gruyter, Inc.
- Martin-Houssart, G., and N. Tabard. 2002. "Représentation Socio-économique du Territoire, Typologie des Quartiers et Communes de France selon la Profession et l'Activité Socio-économique de leurs Habitants." Working Paper F0208. Paris: INSEE.
- Michelis, A.D., M. Estevão, and B.A. Wilson. 2013. "Productivity or Employment: Is it a Choice?" Working Paper 13/97. Washington, DC: IMF.
- Miguel, E., and M.K. Gugerty. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya." *Journal of Public Economics* 89(11–12):2325–68.
- Munshi, K. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *The Quarterly Journal of Economics* 118(2):549–99.
- Nordhaus, W. 2005. "The Sources of the Productivity Rebound and the Manufacturing Employment Puzzle." Working Paper 11354. Cambridge, MA: NBER.
- Ottaviano, G.I., and G. Peri. 2006. "The Economic Value of Cultural Diversity: Evidence from US Cities." *Journal of Economic Geography* 6(1):9–44.
- Ottaviano, G.I.P., and G. Peri. 2012. "Rethinking the Effect of Immigration on Wages." *Journal of the European Economic Association* 10(1):152–97.
- Patacchini, E., and Y. Zenou. 2012. "Ethnic Networks and Employment Outcomes." *Regional Science and Urban Economics* 42(6):938–49.
- Saiz, A. 2007. "Immigration and Housing Rents in American Cities." *Journal of Urban Economics* 61(2):345–71.
- Schmutte, I.M. 2015. "Job Referral Networks and the Determination of Earnings in Local Labor Markets." *Journal of Labor Economics* 33(1):1–32.
- Staiger, D., and J.H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65(3):557–86.
- Wahba, J., and Y. Zenou. 2005. "Density, Social Networks and Job Search Methods: Theory and Application to Egypt." *Journal of Development Economics* 78:443–73.
- Weinberg, B.A., P.B. Reagan, and J.J. Yankow. 2004. "Do Neighborhoods Affect Hours Worked? Evidence from Longitudinal Data." *Journal of Labor Economics* 22(4):891–924.
- Wellman, B. 1996. "Are Personal Communities Local? A Dumptarian Reconsideration." *Social Networks* 18:347–54.
- Winters, P., A. de Janvry, and E. Sadoulet. 2001. "Family and Community Networks in Mexico—U.S. Migration." *Journal of Human Resources* 36(1):159–84.

**D.3 Ethnic Networks and the Informal Labor Market**

# Ethnic Networks and the Informal Labor Market

CAMILLE HÉMET\*

June 2017

## Abstract

This paper develops a dynamic model in which agents belonging to two different groups (majority versus minority) form a network through which they can exchange information about formal or informal job offers. The network plays a central role: both formal and informal job opportunities can only be obtained through word of mouth communication. This network features homophily, so that information circulates imperfectly across groups. This model reveals that workers from the minority face adverse employment conditions on the formal labor market (here due to the reduced size of their network) so that they use undeclared work as an economic safety net. It also shows that lowering informal employment can be achieved through weakening homophily and intensifying the audit rate, although the latter is costly in terms of unemployment.

**Keywords:** Homophily, Networks, Informal Economy

**JEL codes:** A14, J40, K42

---

\*Paris School of Economics

Contact: [camille.hemet@psemail.eu](mailto:camille.hemet@psemail.eu)

Webpage: <http://sites.google.com/site/camillehemet/>

I would like to thank Yann Bramoullé for his guidance and support and Yves Zenou for his advice. I also want to express my gratitude to Birthe Larsen and Ann-Sofie Kolm for insightful discussion at the beginning of this project. This work has also benefited from the comments of participants at the workshop on Shadow Economy, Tax Evasion and Money Laundering (Münster, 2011), at the annual meeting of the European Society for Population Economics (Bern, 2012), at the Econometric Society European Meeting (Malaga, 2012), at the doctoral conference of French Association for the Development of Research in Economics doctoral (Strasbourg, 2013), and at the Louis-André Gérard Varet annual conference (Aix en Provence, 2013), and in particular Nizar Allouch for discussing this paper.

# 1 Introduction

Studies on economic integration and labor market success of immigrants in western economies all reveal that ethnic minorities face adverse labor market conditions. Both [Chiswick \(1980\)](#) and [Borjas \(1994\)](#), working respectively on UK and US data, show that workers from ethnic minorities earn significantly lower wages than majority workers. More recently, [Pendakur and Pendakur \(2011\)](#) report that Canadian-born individuals belonging to visible minorities face a substantial earning gap, which has not eroded over the past twenty years. In a study comparing France, Germany and the UK, [Algan et al. \(2010\)](#) reveal that not only do first and second generation immigrants have lower wages than natives, but they also face higher unemployment rates.<sup>1</sup> In another UK study, [Manacorda et al. \(2012\)](#) even find that recent immigration waves to the UK negatively affect former immigrants' wages. The literature provides various explanations to these low performances, from discrimination ([Becker, 1971](#); [Aeberhardt et al., 2010](#)), spatial mismatch ([Selod and Zenou, 2006](#); [Zenou, 2013](#)), or cultural differences ([Senik and Verdier, 2011](#)), to ethnic identity ([Bisin et al., 2011](#)).

A natural question to ask is how minority workers, especially low-skilled ones, cope with these adverse labor market conditions. The sociological literature argues that one way to compensate for low wages and low labor demand is to enter the informal labor market ([Tienda and Raijman, 2000](#)). [Portes and Haller \(2005\)](#) also highlight the cushioning function of informal employment, that they describe as a common additional source of income when unemployment is high and social benefits meager. In economics as well, [Lemieux et al. \(1994\)](#), using Quebec City data, document that the participation rate of students and unemployed persons in the underground sector amounts to 28 percent, and rises to 32.4 percent for welfare recipients. All those studies suggest that informal economy is a way of improving one's economic condition when facing adverse labor market conditions.

More specifically, I acknowledge that turning to undeclared activities in response to precarious economic situation is particularly relevant for ethnic minorities. Several studies, mostly in sociology, provide empirical support to an over representation of ethnic minorities in the informal economy: in an OECD report, [Lubell \(1991\)](#) gives a few examples illustrating the fact that immigrants are to a large extent part of the informal sector. More recently, [Baganha \(2000\)](#) argues that working in the informal market is the prevailing mode

---

<sup>1</sup>Also refer to [Verdugo \(2009\)](#) for a more detailed description of immigrants labor market situation in France.

of economic integration for immigrants in Portugal, [Light and Gold \(2000\)](#) estimate that 10 percent of the average American ethnic group's workers are employed in the informal sector of ethnic ownership economy, and [Tienda and Raijman \(2000\)](#) find that 38 percent of Mexican immigrant households in Chicago work in the underground economy.

I now turn to more formal arguments supporting the idea that immigrants and ethnic minorities are likely over-represented in the underground economy. First, and as mentioned earlier, minorities face higher unemployment rates and lower wages than the majority group, and hence are more likely to use informal work as an economic safety net. Then, because informal jobs are characterized by the absence of any written contract, and more broadly because they are not submitted to any labor legislation, engaging in informal activities allows to bypass red tape.<sup>2</sup> Given that immigrants are more subject to red tape than natives (e.g. obtaining a work permit), they can be more inclined to turn to informality for this reason as well. In addition to these structural factors, opportunity factors can be brought forward:<sup>3</sup> first, there is evidence that immigrants and ethnic minorities are on average less educated than natives and majority group, and hence are more likely to perform undeclared jobs, that are usually characterized by low-skilled workers ([Boeri and Garibaldi, 2002](#)).<sup>4</sup> In addition, many studies stress the role of networks in the informal economy, showing that ethnic networks are particularly powerful in fostering underground economies.

Alternatively, the extensive participation of ethnic minorities in the informal economy can be explained by what is known in sociology as the *double disadvantage theory*.<sup>5</sup> A standard theory in the literature on ethnic economies, is that economically disadvantaged groups are likely to turn to entrepreneurship and self-employment. This *single disadvantage theory* explains why entrepreneurship rates are so high among ethnic minorities. However, when disadvantaged individuals also lack essential resources to become self-employed (e.g. skills, language, social capital), they turn instead to the informal economy ([Light, 1979](#); [Light and Rosenstein, 1995](#); [Light, 2005](#)).<sup>6</sup>

---

<sup>2</sup>Note that avoiding red tape is one of the main reason for the existence of informal jobs, after tax evasion. See European Commission report on the Eurobarometer pilot survey on undeclared work (2007).

<sup>3</sup>This classification of factors leading to informality is due to [Renooy \(1990\)](#), who distinguishes structural (financial pressure, institutional constraints) and opportunity (skills, education, contacts, environment...) factors.

<sup>4</sup>See [Enste and Schneider \(2000\)](#) and [Gërxhani \(2004\)](#) for comprehensive surveys on the informal economy.

<sup>5</sup>Also known as the *resource-constraint variant* of the disadvantage theory

<sup>6</sup>This quote from [Light and Rosenstein \(1995\)](#) is self-explanatory: "A typical firm in the informal economy requires few resources. For example, fruit vendors at freeway entrances are self-employed, but their informal enterprise did not require extensive resources of money, skill, and knowledge. Anyone could do it. On the other hand, when people with resources confront [labor market] disadvantage, they mobilize those resources to produce a *bona fide* business firm."

From what precedes, one can reasonably think that immigrants and ethnic minorities are more inclined to engage into undeclared activities.<sup>7</sup> To put it another way, current estimates of the labor market situation of ethnic minorities may underestimate their true "activity" rate as well as their income. In this paper, I provide a theoretical background for the fact that ethnic minorities circumvent adverse labor market outcomes by undertaking undeclared jobs. More precisely, this paper develops a dynamic model in which agents belonging to two different ethnic groups randomly meet by pair at each period, and exchange information on two competing activities: formal and informal labor. Formal workers pay taxes, while informal workers evade them. Workers from each group are similar in terms of skills (unskilled workers are more relevant for the considered setting). In this model, networks play a central role, as information about formal and informal job opportunities can only be obtained through word of mouth communication. There is indeed extensive evidence, both in the economic and in the sociological literature that social networks are primordial when it comes to finding a job.<sup>8</sup> The population is characterized by homophily, *i.e* the tendency of individuals to associate with others who are similar to themselves, so that information about job offers circulates imperfectly across groups.

The setting of this paper is close to that of [Calvó-Armengol et al. \(2007\)](#) who study the role of the network and in particular of the strength of interpersonal relationships on the decision to enter criminality versus employment. Yet, the two models differ in three important aspects. First, I consider two different groups instead of a single one in order to introduce some asymmetry in the population in terms of network size. Second, because I want to focus on the role of information transmission, I assume that both formal and informal job offers circulate only through word of mouth communication. By contrast, [Calvó-Armengol et al. \(2007\)](#) suppose that information is diffused through the network for only one of the two activities (crime). Third, I let aside the notion of strong versus weak ties, therefore abandoning the dyadic structure of the network that prevails in their model. This simplifies the model that is already complicated by the first two additional features. Besides, the idea that the relationships are not the same between every two pairs of agents survives through the homophily assumption.

---

<sup>7</sup>Note also that even individuals turning to formal self-employment represent a non negligible share of the informal sector: using data from Denmark, [Kleven et al. \(2011\)](#) show that the tax evasion rate is equal to 41.6% for total self-reported net income, while it is close to zero when income is reported by a third-party.

<sup>8</sup>This evidence is presented in section 2.1 on information transmission.

I analyze the flows of individuals between the various occupations and characterize all the steady-states of this dynamic economy. For this purpose, I solve for the endogenous individual decisions to accept or reject formal and informal job offers. I then derive the conditions under which these steady-states arise in equilibrium. Interestingly, when the net formal wage is larger than the net informal wage, all equilibria involving participation in the informal economy arise for low enough values of the unemployment benefit. This reveals that when the unemployment benefit is too low, individuals are more likely to immediately accept an informal job offer when they receive one rather than to decline and wait for a formal job opportunity.

In particular, an equilibrium in which both groups are formally employed but one also engages in the informal sector arises, and I show that this group is the minority group. This equilibrium is in line with the intuition that the informal sector may be used as an economic safety net when formal employment prospects are low. Indeed, the minority group faces such a low rate of formal job offers arrival (due to the size of the network) that its members are better off immediately accepting an informal job offer than waiting longer for a formal one. On the contrary, unemployed members of the majority group, whose network is more developed and hence who are more likely to receive formal job offers, can afford declining unattractive informal offers.

Simple comparative statics show that when communication across groups is improved, *i.e.* when the degree of homophily of the network decreases, then the formal employment rate increases for both groups, while the informal employment rate is reduced. That is, even if better information transmission can benefit both formal and informal sectors, individuals react by turning down more informal offers. Similarly, I find that increasing the rate at which informal workers are caught by fiscal authorities has a positive impact on the formal employment rate for both groups and a negative effect on the informal employment rate. However, because such a policy does not directly increase the rate of entry toward formal employment, it comes at the expense of increased unemployment rates. On the contrary, reducing homophily has a global negative impact on unemployment.

The remainder of the paper is organized as follows. Section 2 presents the model. In particular, it describes the network structure, information transmission mechanisms and individual incentives. Section 3 defines and characterizes the steady-states, while the conditions for their existence in equilibrium are provided in Section 4. Finally, additional results and some comparative statics are derived in Section 5. Section 6 concludes.

## 2 The model

The population of size  $N$  is composed of two groups, the majority ( $M$ ) and the minority ( $m$ ).  $N_m$  and  $N_M$  respectively denote the size of the minority group and of the majority group in the population, with  $N_m + N_M = N$ , and  $N_M > N_m$ . We also denote  $n_m$  and  $n_M$  the share of each group in the population, such that  $\frac{N_M}{N} + \frac{N_m}{N} = n_M + n_m = 1$ .

The agents are assumed to be homogenous in terms of skills. We can for instance think of low-skilled workers, who are more concerned by undeclared jobs (Boeri and Garibaldi, 2005). The European Commission report on the Eurobarometer pilot survey on undeclared work (2007) also concludes that the main three economic sectors into which respondents admitted to do undeclared work are household services (including housekeeping, gardening, baby-sitting, elderly-sitting), construction services and personnel services. Lemieux et al. (1994) as well report that in Quebec City, workers with a declared job in primary industries have the highest participation rate in the underground sector (17.2 percent), followed by workers with a job in the construction industry (9.4 percent). They also reveal that two thirds of the jobs in the underground economy are in construction and services. Moreover, in response to poor labor market conditions, low-skilled workers are less likely to become self-employed as they lack the necessary resources.

In this model, individuals can engage in three competing activities: declared work, unemployment and undeclared work.<sup>9</sup> There are two main differences between declared and undeclared jobs: undeclared (or informal) jobs are assumed to be more flexible than declared (or formal) jobs,<sup>10</sup> and formal wages are subject to taxation, while undeclared wages are not. However, tax evaders face a positive probability of being detected by fiscal authorities, in which case they have to pay a fine. We denote  $F_{j,t}$  the number of formal workers in group  $j \in \{m, M\}$  at time  $t$ , and  $f_{j,t} \in \frac{F_{j,t}}{N_{j,t}}$  group  $j$ 's formal employment rate at time  $t$ . Similarly,  $U_{j,t}$ , and  $I_{j,t}$ ,  $\forall j \in \{m, M\}$  respectively denote the number of unemployed and informal workers in group  $j \in \{m, M\}$ , while  $u_{j,t}$  and  $i_{j,t}$  stand for group  $j$ 's unemployment and informal employment rates at time  $t$ . They are linked by the following relationship:

$$f_{j,t} + u_{j,t} + i_{j,t} = 1 \quad (1)$$

<sup>9</sup>This model is inspired by Calvó-Armengol et al. (2007), in which agents chose between employment, unemployment and crime.

<sup>10</sup>Section 2.1 provides more details on that issue.

In this economy, agents belong to a network through which they hear about job opportunities, be they formal or informal. There are two reasons why I allow formal job offers to grow through word of mouth communication. First, there is a large and growing literature providing a theoretical basis as well as empirical evidence of the fact that social contacts are a very important channel for finding jobs. I will come back on this literature in section 2.1 on labor market flows, but I refer thereupon to [Ioannides and Datcher Loury \(2004\)](#) and [Topa \(2011\)](#) for extensive surveys of the literature. Second, this paper focuses on the role played by network structure on employment patterns in both regular and underground labor markets. Hence, I do not want to ignore the fact that formal job offers circulate through the network.

## 2.1 Social interactions and job information transmission

**Social interactions.** The pattern of social interactions is close to [Calvó-Armengol et al. \(2007\)](#) or [Montgomery \(1994\)](#), in that individuals randomly meet by pair repeatedly over time. However, while these models are based on the idea that relationships between agents are characterized by their strength ([Granovetter, 1973](#)), I let this notion aside, and abandon the dyadic structure of the network that prevails in their models. That is, agents are not infinitely related to a given other agent who is able to influence their decision. Note however the idea that relationships are not necessarily the same between every two pairs of agents is still there, by assuming ethnic homophily (this will be established below). Therefore, any agent can be matched to an individual from his own ethnic group or from the other group, according to the size of each group in the population and with a bias toward his own group.<sup>11</sup> Similarly, any agent can meet a formal, an informal or an unemployed worker, depending on the various employment rates. Once a pair of agents is randomly formed, the individuals can exchange information about job opportunities. This process is explained below.

**The importance of networks in job information transmission.** The literature offers a plethora of empirical studies supporting the idea that social networks play a key role in finding jobs. As emphasized in [Ioannides and Datcher Loury \(2004\)](#) survey, the first stylized fact emerging from the literature is precisely that workers extensively use social contacts when searching for a job: [Holzer \(1988\)](#) finds that more than 85% of workers use their personal network for job search, while [Corcoran et al. \(1980\)](#) and [Granovetter](#)

---

<sup>11</sup>A possible way to reconcile this simple setting with real world is to think of this population as inhabitants of an ethnically diverse neighborhood: individuals can meet at the bus stop or at the bakery, they identify themselves as neighbors but do not necessarily talk about job opportunities, especially across groups. Another way to consider this set up, in the framework of previous papers, is to say that it focuses on the role of weak ties, *i.e.* random encounters (by opposition to friends and relatives).

(1995) both report that more than 50% of jobs are found through social contacts. More recently, [Bayer et al. \(2008\)](#) estimate that living in the same block increases by more than 33% the probability to work at the same location. Most importantly perhaps in the context of this paper is the fact that finding job through referrals is particularly prominent for low-skilled workers (see [Corcoran et al., 1980](#); [Granovetter, 1995](#); [Topa, 2001](#); [Wahba and Zenou, 2005](#); [Battu et al., 2011](#)).<sup>12</sup> In light of this empirical evidence, I assume without loss of generality that formal jobs opportunities circulate only through the network, and hence I neglect the fact that declared jobs can be obtained through direct formal methods (e.g. employment agencies).<sup>13</sup> Given the illegal nature of undeclared activities, underground job offers cannot be posted by employment agencies or be too widely spread (on the internet for instance) because they incur a risk of sanction by the authorities. Therefore, I also assume that informal jobs opportunities only circulate through the network.

**Transmission of information and labor market flows.** I now turn to a more precise description of the job information transmission mechanisms. As social network is the only channel through which job information circulates, an unemployed worker can hear about a formal (respectively informal) job offer only if s/he is matched with a formal (resp. informal) employed worker who is aware of an offer.<sup>14</sup> Formal workers are aware of a formal job opportunity at an exogenous rate  $\alpha_F$ , and will transmit this information if they meet unemployed workers.<sup>15</sup> Similarly, underground job offers circulate through undeclared workers who hear about an opportunity at an exogenous rate  $\alpha_I$ , and who transmit it to unemployed workers if they are randomly matched to one. I assume that the informal sector is more flexible than the formal one, so that  $\alpha_I > \alpha_F$ . As already mentioned, I assume that this economy features homophily, that is, individuals tend to form links and gather with people similar to themselves, and in particular along ethnic lines (see [McPherson et al. \(2001\)](#) for a review of the literature on homophily, and [Currarini et al. \(2009\)](#) and [Bramoullé et al. \(2012\)](#) for model of network formation featuring homophily). As a consequence, (formal or informal) employed workers will systematically transmit information to individuals of their own ethnic group, but they will communicate a job offer to someone of the other ethnic group with a probability  $\beta \in (0, 1)$ . In other words, the network is characterized by a bias making it more difficult to get information from someone of a

---

<sup>12</sup>Also see [Montgomery \(1991\)](#), [Calvó-Armengol and Jackson \(2004\)](#) and [Galenianos \(2010\)](#) for theoretical work on the role of networks in the labor market.

<sup>13</sup>By contrast, [Calvó-Armengol et al. \(2007\)](#) assume that jobs can only be found through direct formal methods and that the information going through social contacts concerns crime opportunities only.

<sup>14</sup>The general idea is that workers are aware of job creations or vacancies in their own company.

<sup>15</sup>There is no on-the-job search in this model.

different group (*inbreeding bias*), or to meet someone (*meeting bias*) of the other ethnic group.<sup>16</sup> As emphasized by [Fearon and Laitin \(1996\)](#), "ethnic groups are often characterized by relatively dense social networks and low-cost access to information". [Bertrand et al. \(2000\)](#) also provide empirical evidence of strong information transmission (about welfare benefits) within language group, at the neighborhood level, supporting the assumption that information circulate better within than across ethnic groups. We can also interpret  $\beta$  as an inverse measure of the degree of spatial segregation between the minority and the majority groups (or as the degree of integration of the minority group): the lower  $\beta$ , the higher the segregation, hence the lower the probability to find a job through someone from a different ethnic group.<sup>17</sup>

I finally close the description of labor market flows with job destruction, which is assumed exogenous in both sectors. I denote  $\delta_F$  the rate at which formal workers lose their job, and  $\delta_I$  the rate at which an undeclared worker is fired. I assume that  $\delta_I > \delta_F$  as the informal worker is not protected by any job contract, making it easier to lay him off. In addition, informal jobs can also be terminated if they are audited by fiscal authorities, which happens at a rate  $p$ . In the end, informal workers lose their job at rate  $\delta_I + p$ .

## 2.2 Flows of dyads between states

From what precedes, I characterize the net flows of the formal employment, unemployment and informal employment rates between  $t$  and  $t + \Delta t$  for each group  $j \in \{m, M\}$  as follows.

$$\begin{cases} \dot{F}_{j,t} = U_{j,t} \left( \frac{F_{j,t}}{N} + \beta \frac{F_{-j,t}}{N} \right) \alpha_F \phi_j - \delta_F F_{j,t} \\ \dot{U}_{j,t} = \delta_F F_{j,t} + (\delta_I + p) I_{j,t} - U_{j,t} \left[ \left( \frac{F_{j,t}}{N} + \beta \frac{F_{-j,t}}{N} \right) \alpha_F \phi_j + \left( \frac{I_{j,t}}{N} + \beta \frac{I_{-j,t}}{N} \right) \alpha_I \psi_{j,t} \right] \\ \dot{I}_{j,t} = U_{j,t} \left( \frac{I_{j,t}}{N} + \beta \frac{I_{-j,t}}{N} \right) \alpha_I \psi_j - (\delta_I + p) I_{j,t} \end{cases} \quad (2)$$

Let us explain the first equation, to get the general intuition for this system:  $\dot{F}_{j,t}$  is the change in the number of formally employed workers between  $t$  and  $t + \Delta t$ . At each point in time, formal jobs are destroyed at rate  $\delta_F$ . On the contrary, a formal job is created when an unemployed worker hears of a formal job

<sup>16</sup>Montgomery (1991) also assumes inbreeding bias, but along skills rather than ethnicity.

<sup>17</sup>Zenou (2013) develops a similar model with two ethnic groups and one activity, embedded in a urban framework. He also introduces separation, but in a different manner: spatial segregation implies that agents further away from the business district meet less weak ties, whether belonging to the majority or to the minority group.

opportunity. This happens if he is matched with a formal worker from any ethnic group (there is a probability  $\frac{F_j}{N}$  to meet a formal worker of group  $j$ ), who is aware of a job offer (which happens at rate  $\alpha_F$ ), and who transmits the information (with probability 1 if they belong to the same ethnic group, and  $\beta$  otherwise due to the in-group bias).<sup>18</sup>

In addition, once an unemployed worker hears of a formal job offer, he may still decide not to take the job and stay unemployed. This is embodied by the choice variable  $\phi_j$ , which is a dummy variable equal to one if the unemployed agent decides to accept the formal job offer, and equal to zero otherwise. Similarly, in the third equation,  $\psi_j$  is the dummy variable representing the decision to accept an undeclared job opportunity. These decision variables will be endogenized afterwards, when individuals' incentives are introduced.

Noting that  $\frac{F_{j,t}}{N} = \frac{F_{j,t}}{N} \frac{N_j}{N_j} = f_{j,t} n_j$ , and dividing both sides of each equation by  $N_j$ , system (2) can be rewritten as follows.

$$\begin{cases} \dot{f}_{j,t} = u_{j,t}(f_{j,t} n_j + \beta f_{-j,t} n_{-j}) \alpha_F \phi_j - \delta_F f_{j,t} \\ u_{j,t} = \delta_F f_{j,t} + (\delta_I + p) i_{j,t} - u_{j,t} [(f_{j,t} n_j + \beta f_{-j,t} n_{-j}) \alpha_F \phi_j + (i_{j,t} n_j + \beta i_{-j,t} n_{-j}) \alpha_I \psi_j] \\ \dot{i}_{j,t} = u_{j,t}(i_{j,t} n_j + \beta i_{-j,t} n_{-j}) \alpha_I \psi_j - (\delta_I + p) i_{j,t} \end{cases} \quad (3)$$

Finally, using equation (1), this system reduces to the following two dynamic equations system.

$$\begin{cases} \dot{f}_{j,t} = u_{j,t}(f_{j,t} n_j + \beta f_{-j,t} n_{-j}) \alpha_F \phi_j - \delta_F f_{j,t} \\ \dot{i}_{j,t} = u_{j,t}(i_{j,t} n_j + \beta i_{-j,t} n_{-j}) \alpha_I \psi_j - (\delta_I + p) i_{j,t} \end{cases} \quad (4)$$

### 2.3 Incentives

As mentioned earlier, the choice variables  $\phi_j$  and  $\psi_j$  are endogenously determined taking into account the incentives faced by individuals in their decision making. Declared workers earn a wage  $w_F$ , and pay a unit tax  $t$  so that their net wage is  $w_F(1 - t)$ . Unemployed workers receive a financial payoff  $b$ , which can be

<sup>18</sup>I can also put it in the following way: he has a  $\frac{N_j}{N} f_{t_j}$  probability to meet a formal worker of his own group and a  $\beta \frac{N_{-j}}{N} f_{t_{-j}}$  probability to meet one of the other ethnic group. In this case, I can also view  $\beta$  as a *meeting* bias, as in Currarini et al. (2009). This has a slightly different interpretation: in the case of the *inbreeding* bias,  $\beta$  means that one is less willing to transmit information to a member of a different group, while in the case of the *meeting* bias,  $\beta$  means that one is less likely to meet a member of the other ethnic group. In both cases, the consequence is that information circulates better within than across ethnic groups.

considered as the unemployment insurance, as social welfare or as the value of leisure.<sup>19</sup> I further assume that the net formal wage is larger than the unemployment benefit:  $w_F(1-t) > b$ . Finally, undeclared workers earn a wage  $w_I$ , and evade from paying taxes. However, they can be detected by fiscal authorities with probability  $p$ , which implies paying a fine  $f$  defined as a share of the wage perceived. Therefore, the expected informal wage is given by:  $w_I(1-pf)$ .

Agents make their decisions on the basis of their future employment prospects. In the long-run, individual values for each possible labor market status are given by the following Bellman equations, where  $V_j^S$  denotes the expected lifetime utility of a member of group  $j \in \{m, M\}$  involved in sector  $S \in \{F, I, U\}$ .

$$rV_j^F = w_F(1-t) + \delta_F(V_j^U - V_j^F) \quad (5)$$

$$rV_j^U = b + (f_{j,t}n_j + \beta f_{-j,t}n_{-j})\alpha_F\phi_j(V_j^F - V_j^U) + (i_{j,t}n_j + \beta i_{-j,t}n_{-j})\alpha_I\psi_j(V_j^I - V_j^U) \quad (6)$$

$$rV_j^I = w_I(1-pf) + (\delta_I + p)(V_j^U - V_j^I) \quad (7)$$

In these equations,  $r$  denotes the interest rate. Equation (5) tells that the lifetime expected utility of a formally employed worker is equal to his wage, plus the net value of losing his job (*i.e.* the value of being unemployed compared to the value of being in a formal job). The other equations have a similar interpretation.

### 3 Steady-state analysis

At a steady-state ( $f_j^*$ ,  $u_j^*$ ,  $i_j^*$ ) the number of workers in each sector (and by extension the various employment rates) is stable, (*i.e.* the net flows in each employment status is equal to zero), and I get the following relationships using the set of dynamic equations (4).

$$f_j^* = \frac{(f_j^*n_j + \beta f_{-j}^*n_{-j})\alpha_F\phi_j}{\delta_F} u_j^* \quad (8)$$

---

<sup>19</sup>The last two may be a better interpretation of  $b$ , as individuals are unlikely to receive unemployment insurance after performing undeclared work.

$$i_j^* = \frac{(i_j^* n_j + \beta i_{-j}^* n_{-j}) \alpha_I \psi_j}{\delta_I + p} u_j^* \quad (9)$$

$$u_j^* = 1 - f_j^* - i_j^* \quad (10)$$

There exist sixteen different types of steady-states depending on the values of the formal and informal employment rates (e.g. null or strictly positive). However, they can be reduced to the following ten steady-states by symmetry between  $j$  and  $-j$ ,  $\forall j \in \{m, M\}$ :

(i)	Full unemployment	$f_j^* = f_{-j}^* = 0$	$i_j^* = i_{-j}^* = 0$
(ii)	One group in the formal sector only	$f_j^* > 0, f_{-j}^* = 0$	$i_j^* = i_{-j}^* = 0$
(iii)	Both groups in the formal sector only	$f_j^* > 0, f_{-j}^* > 0$	$i_j^* = 0, i_{-j}^* = 0$
(iv)	One group in the informal sector only	$f_j^* = f_{-j}^* = 0$	$i_j^* > 0, i_{-j}^* = 0$
(v)	One group in formal and informal sectors	$f_j^* > 0, f_{-j}^* = 0$	$i_j^* > 0, i_{-j}^* = 0$
(vi)	One group in each sector	$f_j^* = 0, f_{-j}^* > 0$	$i_j^* > 0, i_{-j}^* = 0$
(vii)	Both groups in the formal, only one in the informal	$f_j^* > 0, f_{-j}^* > 0$	$i_j^* > 0, i_{-j}^* = 0$
(viii)	Both groups in the informal sector only	$f_j^* = f_{-j}^* = 0$	$i_j^* > 0, i_{-j}^* > 0$
(ix)	Both groups in the informal, only one in the formal	$f_j^* > 0, f_{-j}^* = 0$	$i_j^* > 0, i_{-j}^* > 0$
(x)	Both groups in both sectors	$f_j^* > 0, f_{-j}^* > 0$	$i_j^* > 0, i_{-j}^* > 0$

I now characterize each of the ten steady-states presented above: I indicate the values of the choice variables  $\phi_j$  and  $\varphi_j$  compatible with each steady-state and derive expressions of the steady-state employment rates. Appendix A proves the results presented in this section.

(i) A full unemployment steady-state ( $f_j^* = f_{-j}^* = 0$  and  $i_j^* = i_{-j}^* = 0$ ) when  $\phi_j, \phi_{-j}, \psi_j$  and  $\psi_{-j} \in \{0, 1\}$ .

(ii) Steady-states where one group is fully unemployed, while the other takes part in the formal economy

only ( $f_j^* > 0, f_{-j}^* = 0$  and  $i_j^* = i_{-j}^* = 0$ ) when  $\phi_j = 1, \phi_{-j} = 0$  and  $\psi_j, \psi_{-j} \in \{0, 1\}$ . In this case,

$$f_j^* = 1 - \frac{1}{n_j} \frac{\delta_F}{\alpha_F}.^{20}$$

<sup>20</sup>Obviously, because job offers circulate through the network, the larger the size of the group  $n_j$ , the larger  $f_j^*$ . Note then that if the group that has a positive formal employment rate ( $j$ ) is the minority group, its formal employment rate is always smaller than the formal employment rate of the majority group in the case where the majority group would be the one to have a positive formal employment rate.

**(iii)** A no-informal economy steady-state ( $f_j^* > 0$ ,  $f_{-j}^* > 0$ ,  $i_j^* = 0$  and  $i_{-j}^* = 0$ ) when  $\phi_j = \phi_{-j} = 1$  and  $\psi_j, \psi_{-j} \in \{0, 1\}$ ,  $\forall j$ . In this case, the equilibrium values of the formal employment rates are characterized by  $f_j^* = \frac{(f_j^* n_j + \beta f_{-j}^* n_{-j}) \alpha_F}{\delta_F + (f_j^* n_j + \beta f_{-j}^* n_{-j}) \alpha_F} < 1$ ,  $\forall j \in \{m, M\}$ .

**(iv)** Steady-states where one group is fully unemployed, while the other takes part in the informal economy only ( $f_j^* = f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ) when  $\phi_j, \phi_{-j} \in \{0, 1\}$ ,  $\psi_j = 1$  and  $\psi_{-j} = 0$ . In this case,  $i_j^* = 1 - \frac{1}{n_j} \frac{\delta_I + p}{\alpha_I}$ .

**(v)** Steady-states where one group is fully unemployed while the other takes part in both employment sectors ( $f_j^* > 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ) when  $\phi_j = 1$ ,  $\phi_{-j} = 0$ ,  $\psi_j = 1$  and  $\psi_{-j} = 0$ . In this steady state,  $f_j^*$  and  $i_j^*$  are characterized by  $1 - f_j^* - i_j^* = \frac{1}{n_j} \frac{\delta_F}{\alpha_F}$  and  $1 - f_j^* - i_j^* = \frac{1}{n_j} \frac{\delta_I + p}{\alpha_I}$ , implying that the relative destruction rate is the same in both sectors ( $\frac{\delta_F}{\alpha_F} = \frac{\delta_I + p}{\alpha_I}$ ).

**(vi)** Steady-states where one group works in the informal sector only while the other takes part in the formal labor market only ( $f_j^* = 0$ ,  $f_{-j}^* > 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ) when  $\phi_j = 0$ ,  $\phi_{-j} = 1$ ,  $\psi_j = 1$ ,  $\psi_{-j} = 0$ . In this case, we have the following values:

$$f_j^* = 0, i_j^* = 1 - \frac{1}{n_j} \frac{\delta_I + p}{\alpha_I} \text{ and } u_j^* = \frac{1}{n_j} \frac{\delta_I + p}{\alpha_I}$$

$$f_{-j}^* = 1 - \frac{1}{n_{-j}} \frac{\delta_F}{\alpha_F}, i_{-j}^* = 0 \text{ and } u_{-j}^* = \frac{1}{n_{-j}} \frac{\delta_F}{\alpha_F}$$

**(vii)** Steady-states where one group is involved in both sectors, while the other works in the formal sector only ( $f_j^* > 0$ ,  $f_{-j}^* > 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ) when  $\phi_j = \phi_{-j} = 1$ ,  $\psi_j = 1$ , and  $\psi_{-j} = 0$ . In this case, the employment rates are determined as follows.

$$f_j^* = \beta \left[ \frac{n_{-j}}{n_j} \frac{\alpha_F (\delta_I + p)}{\alpha_I \delta_F - \alpha_F (\delta_I + p)} - \frac{1}{n_j} \frac{\delta_F (\delta_I + p)}{\alpha_I \delta_F - (1 - \beta^2) \alpha_F (\delta_I + p)} \right] \quad (11)$$

$$f_{-j}^* = 1 - \frac{1}{n_{-j}} \frac{\delta_F}{\alpha_F} \frac{\alpha_I \delta_F - \alpha_F (\delta_I + p)}{\alpha_I \delta_F - (1 - \beta^2) \alpha_F (\delta_I + p)} \quad (12)$$

$$i_j^* = 1 - \frac{1}{n_j} \frac{\delta_I + p}{\alpha_I} - f_j^* \quad (13)$$

**(viii)** An informal economy steady-state ( $f_j^* = f_{-j}^* = 0$  and  $i_j^* > 0$ ,  $i_{-j}^* > 0$ ) when  $\phi_j, \phi_{-j} \in \{0, 1\}$ ,  $\psi_j = 1$

and  $\psi_{-j} = 1$ . In this case, the equilibrium values of the informal employment rates are characterized by  $i_j^* = \frac{(i_j^* n_j + \beta i_{-j}^* n_{-j}) \alpha_I}{\delta_I + p + (i_j^* n_j + \beta i_{-j}^* n_{-j}) \alpha_I} < 1, \forall j \in \{m, M\}$ .

(ix) Steady-states where one group is involved in both sectors, while the other works in the informal sector only ( $f_j^* > 0, i_j^* > 0$  and  $f_{-j}^* = 0, i_{-j}^* > 0$ ) when  $\phi_j = 1, \phi_{-j} = 0, \psi_j = 1$  and  $\psi_{-j} = 1$ . In this case, the employment rates are determined as follows.

$$i_j^* = \beta \left[ \frac{n_{-j}}{n_j} \frac{\alpha_I \delta_F}{\alpha_F (\delta_I + p) - \alpha_I \delta_F} - \frac{1}{n_j} \frac{\delta_F (\delta_I + p)}{\alpha_F (\delta_I + p) - (1 - \beta^2) \alpha_I \delta_F} \right]$$

$$i_{-j}^* = 1 - \frac{1}{n_{-j}} \frac{\delta_I + p}{\alpha_I} \frac{\alpha_F (\delta_I + p) - \alpha_I \delta_F}{\alpha_F (\delta_I + p) - (1 - \beta^2) \alpha_I \delta_F}$$

$$f_j^* = 1 - \frac{1}{n_j} \frac{\delta_F}{\alpha_F} - i_j^*$$

(x) A mixed steady-state ( $f_j^* > 0, f_{-j}^* > 0, i_j^* > 0$  and  $i_{-j}^* > 0$ ) when  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ . In this case, the employment rates are characterized as follows,  $\forall j$ .

$$f_j^* = \frac{(f_j^* n_j + \beta f_{-j}^* n_{-j}) \alpha_F}{\delta_F} (1 - f_j^* - i_j^*)$$

$$i_j^* = \frac{(i_j^* n_j + \beta i_{-j}^* n_{-j}) \alpha_I}{\delta_I + p} (1 - f_j^* - i_j^*)$$

However, some of these steady-states never arise in equilibrium. The next section is devoted to define and characterize the various equilibria. In particular, I derive the conditions under which the steady-states presented above arise in equilibrium.

## 4 Equilibria

### 4.1 Unemployed workers decisions

The conditions of existence of the various equilibria are derived from the conditions determining the value of the two choice variables  $\phi_j$  and  $\psi_j$ , *i.e.* the conditions under which unemployed workers are willing to accept or reject a formal or an informal job offer. To do so, I simply compare the lifetime expected utility of an unemployed worker to the one he would get from accepting a formal or an informal job offer. Combining

equations (5), (6) and (7), I get:

$$V_j^U - V_j^F = \frac{[b - w_f(1 - t)][r + \delta_I + p + (i_j n_j + \beta i_{-j} n_{-j})\alpha_I \psi_j] + [w_I(1 - pf) - b][(i_j n_j + \beta i_{-j} n_{-j})\alpha_I \psi_j]}{(r + \delta_F)[r + \delta_I + p + (i_j n_j + \beta i_{-j} n_{-j})\alpha_I \psi_j] + (r + \delta_I + p)[(f_j n_j + \beta f_{-j} n_{-j})\alpha_F \phi_j]} \quad (14)$$

and

$$V_j^I - V_j^U = \frac{[w_I(1 - pf) - b][r + \delta_F + p + (f_j n_j + \beta f_{-j} n_{-j})\alpha_F \phi_j] + [b - w_f(1 - t)][(f_j n_j + \beta f_{-j} n_{-j})\alpha_F \phi_j]}{(r + \delta_F)[r + \delta_I + p + (i_j n_j + \beta i_{-j} n_{-j})\alpha_I \psi_j] + (r + \delta_I + p)[(f_j n_j + \beta f_{-j} n_{-j})\alpha_F \phi_j]} \quad (15)$$

Equation (14) helps determine the conditions under which an unemployed agent will decide to accept a formal job offer: he will accept such an offer if the lifetime expected value of this declared job is larger than the lifetime expected value of unemployment, *i.e.* if  $V_j^U - V_j^F \leq 0$ . Similarly, an unemployed worker will decide to accept an informal job offer when  $V_j^I - V_j^U \geq 0$ . Equation (15) is useful to find the conditions under which this inequality holds. More precisely, after noting that the denominator of the above equations is always positive and re-arranging the numerators, we can derive the following conditions.

$$\phi_j = 1 \Leftrightarrow V_j^F \geq V_j^U \Leftrightarrow (r + \delta_I + p)[w_F(1 - t) - b] \geq [w_I(1 - pf) - w_F(1 - t)](i_j n_j + \beta i_{-j} n_{-j})\alpha_I \psi_j \quad (16)$$

and

$$\psi_j = 1 \Leftrightarrow V_j^I \geq V_j^U \Leftrightarrow (r + \delta_F)[w_I(1 - pf) - b] \geq [w_F(1 - t) - w_I(1 - pf)](f_j n_j + \beta f_{-j} n_{-j})\alpha_F \phi_j \quad (17)$$

Both results are rather intuitive: from equation (16), we see that an unemployed worker will accept a formal job offer if the net financial gain from becoming a declared worker is large enough compared to the informal wage net of the formal wage. Similarly, (17) tells us that for an unemployed worker to accept an informal job offer, the net financial gain from becoming an undeclared worker must be large enough compared to the formal wage net of the informal wage. Otherwise stated, an unemployed agent compares his immediate net gain (left hand side terms) to the gain he would make in the outside option, would he refuse the offer in one sector to wait for an offer in the other sector (right hand side terms).

## 4.2 Equilibria

I now examine the conditions under which the various steady-states arise in equilibrium. The results are proved in Appendix B. Before taking a closer look at the equilibria, it is important to underline that four

steady-states never arise in equilibrium:

- ◊ Steady-states (ii) where one group is fully unemployed, while the other takes part in the formal economy only ( $f_j^* > 0$ ,  $f_{-j}^* = 0$  and  $i_j^* = i_{-j}^* = 0$ ,  $\forall j$ ),
- ◊ Steady-states (iv) where one group is fully unemployed, while the other takes part in the informal economy only ( $f_j^* = f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j$ ),
- ◊ Steady-states (v) where one group is fully unemployed while the other takes part in both employment sectors ( $f_j^* > 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j$ ),
- ◊ Steady-states (vi) where one group works in the informal sector only while the other takes part in the formal labor market only ( $f_j^* = 0$ ,  $f_{-j}^* > 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j$ ).

The first three cases correspond to situations where one group takes part in at least one of the two employment sectors, while the other group is fully unemployed. The reason why this type of situation cannot happen is very intuitive and related to the information transmission mechanisms. Because there is always at least some information transmission across groups (except in the very extreme case where  $\beta = 0$ ), it is not possible that a group never hears about any job offer (and hence end up fully unemployed) when part of the other group is occupied.<sup>21</sup> The intuition for the last case (vi) not arising in equilibrium is a bit less direct, and is related to incentives in addition to information transmission mechanisms. The idea is that both groups have similar incentives, as they face the same wages and the same exogenous transition rates between employment status. So there is no reason why a group would engage in formal work while the other would go toward informality. Let us now turn to the equilibria. We have two different sets of equilibria arising according to the value of the relative net wage.

---

<sup>21</sup>In an earlier version of this model, I assume that individuals from different ethnic group do not share informal job information, *i.e.* that  $\beta = 0$  in the informal sector. The idea behind this assumption was the following. Because undeclared workers act beyond the law and face a risk of detection and punishment for evading taxes, they do not want to spread this kind of information too widely. Otherwise stated, they share undeclared job opportunities with people they really trust, and I assumed this could not happen outside one's own ethnic group. With this assumption, steady-state (iv) could arise in equilibrium.

◇ **Suppose first that**  $w_F(1-t) > w_I(1-pf)$ :

- The full unemployment steady-state (i), *i.e.*  $f_j^* = 0$  and  $i_j^* = 0, \forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_F(1-t)$ , that is always.
- The no informal sector steady-state (iii), *i.e.*  $f_j^* > 0$  and  $i_j^* = 0, \forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_F(1-t)$ , that is always.
- Steady-states (vii) where one group is involved in both sectors, while the other works in the formal sector only, *i.e.*  $f_j^* > 0, i_j^* > 0, f_{-j}^* > 0$  and  $i_{-j}^* = 0, \forall j \in \{m, M\}$  arises in equilibrium as long as

$$\begin{aligned} w_I(1-pf) - [w_F(1-t) - w_I(1-pf)](n_{-j}f_{-j} + \beta n_j f_j) \frac{\alpha_F}{r + \delta_F} < b \\ \leq w_I(1-pf) - [w_F(1-t) - w_I(1-pf)](n_j f_j + \beta n_{-j} f_{-j}) \frac{\alpha_F}{r + \delta_F} \end{aligned}$$

with  $j$  and  $-j$  such that  $n_j f_j < n_{-j} f_{-j}$ .

- The no formal sector steady-state (viii), *i.e.*  $f_j^* = 0$  and  $i_j^* > 0, \forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_I(1-pf)$ .
- The mixed steady-state (x), *i.e.*  $f_j^* > 0$  and  $i_j^* > 0, \forall j \in \{m, M\}$  arises in equilibrium if

$$b \leq w_I(1-pf) - [w_F(1-t) - w_I(1-pf)](n_{-j}f_{-j} + \beta n_j f_j) \frac{\alpha_F}{r + \delta_F}$$

with  $j$  and  $-j$  is such that  $n_j f_j < n_{-j} f_{-j}$ .

◇ **Suppose now that**  $w_F(1-t) \leq w_I(1-pf)$ :

- The full unemployment steady-state (i), where  $f_j^* = 0$  and  $i_j^* = 0, \forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_F(1-t)$ , that is always.
- The no informal sector steady-state (iii), where  $f_j^* > 0$  and  $i_j^* = 0, \forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_F(1-t)$ , that is always.
- The no formal sector steady-state (viii), *i.e.*  $f_j^* = 0$  and  $i_j^* > 0, \forall j \in \{m, M\}$  arises in equilibrium if  $b \leq w_F(1-t)$ , that is always.
- Steady-states (ix) where one group is involved in both sectors, while the other works in the informal

sector only, *i.e.*  $f_j^* > 0$ ,  $i_j^* > 0$ ,  $f_{-j}^* = 0$  and  $i_{-j}^* > 0$ ,  $\forall j \in \{m, M\}$  arises in equilibrium as long as

$$\begin{aligned} w_F(1-t) - [w_I(1-pf) - w_F(1-t)](n_{-j}i_{-j} + \beta n_j i_j) \frac{\alpha_I}{r+\delta_I+p} &< b \\ &\leq w_F(1-t) - [w_I(1-pf) - w_F(1-t)](n_j i_j + \beta n_{-j} i_{-j}) \frac{\alpha_I}{r+\delta_I+p} \end{aligned}$$

with  $j$  and  $-j$  are such that  $n_j i_j < n_{-j} i_{-j}$ .

### 4.3 Intuitions and comments

A few comments are in order here. We can first observe that the full-unemployment steady-state always arises in equilibrium, whatever the relative net wage. This feature directly comes from the set up of the model, and is more precisely due to the simplifying assumption that job offers go only through the network. Under this assumption, it is indeed impossible that an unemployed worker hears about a job offer when there are no workers in the economy, so that a situation where no agent is employed in any sector is sustainable. Note that the same argument applies to equilibria (iii) and (viii) in which the employment rate in a given sector is null for both groups. Of course, if the model allowed for part of the job offers to arrive exogenously and independently of the network -as is the case in [Calvó-Armengol et al. \(2007\)](#)- this steady-state would never be an equilibrium.

Interestingly, the no-informal steady-state (iii) always arises in equilibrium, whether the net formal wage is larger than the net informal wage or not. The former situation is easy to understand: the formal wage is attractive enough to make unemployed workers systematically accept formal job offers when they find one. In addition, because job offers only circulate through word of mouth communication, the absence of workers in the informal sector is sustainable as explained above. The case where the informal wage is larger holds for the same reasons, even though slightly less intuitive: the absence of informal workers prevents informal job finding, and workers always accept formal job offers which are still more attractive than unemployment. However, this situation is somehow realistic as one could hardly assume the existence of exogenous informal job offers (e.g. employment agencies or adverts).

Let me now jump to the no-formal-sector steady-state (viii), which is symmetric to case (iii) in terms of activity. As expected, it always arises as an equilibrium when the net wage is larger in the informal sector,

for the same reasons that (iii) always arises when the net wage is larger in the formal sector (*i.e.* appropriate financial incentives combined with the lack of outside options due to the absence of worker in the other sector). Yet, when the formal wage is larger than the informal one, the no-formal-sector equilibrium only exists if  $b$  is sufficiently low, and more precisely lower than the informal wage  $w_I(1 - pf)$ . The intuition is that with such a low unemployment benefit, individuals prefer accepting an informal job offer rather than staying unemployed and waiting for a better job offer. It is even trivial in this peculiar equilibrium where more financially attractive offers (formal job offers) are never to arrive.<sup>22</sup>

Let me now turn to the asymmetric equilibrium (vii) in which one group is engaged in both sectors while the other specializes in the formal sector.<sup>23</sup> The first thing to note is that this equilibrium arises only when the formal wage is larger than the informal wage. Thus, even if the informal sector is less attractive in terms of (net) wages than the formal sector, one of the two groups engages into it. This equilibrium therefore illustrates the intuition that the informal sector may be used as an economic safety net when formal employment prospects are low.

In particular, this equilibrium arises for a given range of  $b$ , which is such that  $b < w_I(1 - pf)$ .<sup>24</sup> Therefore, because the unemployment benefit is so low, some unemployed individuals are more likely to accept an informal job offer when they receive one (even if they would rather turn to the formal sector) instead of staying unemployed longer. However, even if the wage differential is the same for both groups, we know from Section 3 that they do not take the same decision *vis-à-vis* informal job offers: group  $j$  always accepts ( $\psi_j = 1$ ), group  $-j$  always declines ( $\psi_{-j} = 0$ ). This divergent behavior can be explained by differences in the relative rate of access to a given sector, which depends on the job opening and job destruction rates ( $\alpha_S$  and  $\delta_S$ ,  $S \in \{F, I\}$ ), but also on the size of the network. Re-arranging relation (17) indicates that a group  $j$  individual accepts an informal job offer if  $[w_I(1 - pf) - b] \geq [w_F(1 - t) - w_I(1 - pf)](f_j n_j + \beta f_{-j} n_{-j}) \frac{\alpha_F}{r + \delta_F}$ , that is if the present net gain from accepting this offer is larger than the expected net gain from formal employment, would he wait for a formal offer instead. So, in this equilibrium, group  $j$ 's probability to receive a formal job offer is so low that its members always accept informal job offers, while group  $-j$

<sup>22</sup>This idea matches the notion of unemployment as a "waiting room" developed in Calvó-Armengol et al. (2007).

<sup>23</sup>Equilibrium (ix) in which one group is engaged in both sectors while the other specializes in the informal sector perfectly mirrors case (vii) in terms of sectors. Thus, all the comments made for (vii) apply by symmetry to (ix).

<sup>24</sup>It is direct to show that the upper bound for  $b$  displayed in subsection (4.2) is strictly lower than  $w_I(1 - pf)$  when  $w_F(1 - t) > w_I(1 - pf)$ .

members receive formal job offers frequently enough to afford declining an informal offer and wait for a formal one.

Crucially, I show that in this equilibrium, the group that engages into the informal sector is always the minority group. This is proved in Appendix C. Once again, this is in line with the informal sector acting as an economic safety net for individuals belonging to the minority: because their network is less developed, they face adverse labor market conditions, that is they have a low probability of finding a formal job so that they will accept informal job offers in order to avoid staying unemployed.

I finally want to insist on the role of the unemployment benefit in defining the equilibria. Let us consider the case where the formal wage is larger than the informal wage. We can see that all the equilibria in which at least one group engages in the informal sector ((vii), (viii) and (x)) are equilibria in which the unemployment benefit is under a given threshold. More generally, all the conditions under which these equilibria emerge are such that  $b \leq w_I(1 - pf)$ . As explained above, in this situation unemployed workers are more likely to accept any job offer, even an informal one. Therefore, if unemployment benefits were larger, unemployed workers incentives would be affected such that they could afford declining informal job offers and wait until they hear about a formal one (*i.e.* switch to a different equilibrium).

## 5 Further results and comparative statics

In this section, I look more precisely at the equilibria characteristics in terms of group composition, focusing on the case where the net formal wage is larger than the net informal wage. The alternative case is not considered here, but most of the results and similar reasoning can be applied by symmetry. I also let the trivial full unemployment equilibrium (i) aside, and concentrate on equilibria (iii) and (vii).

An important feature of equilibrium (iii) is that the majority group always has a larger employment rate than the minority group ( $f_M^* \geq f_m^*$ ), due to the size of the network. To put it differently, the minority group always faces a larger unemployment rate. A formal proof of this result is given in Appendix D. The intuition is simply that the wider one's network, the larger one's probability to meet another individual, hence the larger the probability to hear about a job offer.<sup>25</sup> Because the network features homophily, the transmission

<sup>25</sup>This result is consistent with one of those in Zenou (2013) (proposition 9).

of information is imperfect across groups, and hence members of the majority group always have access to a larger network than the minority group.

A subsequent intuition is that the employment rates should increase with the degree of integration between the two groups, *i.e.* be larger for smaller values of  $\beta$ . This equilibrium is quite difficult to solve analytically, as  $f_j^*$  and  $f_{-j}^*$  are determined by two equations of two unknowns, and are the roots of third order polynomials (see Appendix A for more details). Yet, it is easy and informative to look at the two extreme sub-cases of complete segregation ( $\beta = 0$ ) and no segregation ( $\beta = 1$ ). When  $\beta = 0$ , *i.e.* when there is no transmission of information across groups, we obtain  $f_m^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{n_m}$  and  $f_M^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{n_M}$ . Obviously,  $f_M^* > f_m^*$ . As job information can only come from one's own group,  $f_j$  is directly proportional to  $n_j$ . When  $\beta = 1$ , *i.e.* when information circulates perfectly between ethnic groups, we get  $f_m^* = f_M^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{(n_m + n_M)} = 1 - \frac{\delta_F}{\alpha_F}$ . The equality between the employment rates of the two groups is very intuitive as information circulates perfectly across groups so that both groups form a unique large group. Note that the no-segregation employment rate is larger than the full-segregation employment rate for both groups, because each individual not only has access to information from his own group, but also from the other one. More importantly, the increase in the employment rate between the full segregation and the no segregation cases is larger for the minority group. This is due to the fact that when segregation decreases, the new information network to which the minority group has access is relatively larger than that of the majority group (which has already access to a large information network in the full-segregation case). In other words, the increase in the size of the majority group information network is marginal. Finally, these two simple cases reveal that the employment rates decrease with the relative exit rate out of employment  $\frac{\delta_F}{\alpha_F}$ , as foreseeable.<sup>26</sup>

Let me now turn to equilibrium (vii), in which one group specializes in the formal sector, while the other has to engage in the informal sector as a result of low formal employment prospects. This equilibrium is more convenient to study: the various employment rates can be analytically solved for, as shown in Section 3. Using equations (11) to (13), I can show that:

$$\frac{\partial f_{-j}^*}{\partial \beta} > 0, \quad \frac{\partial f_j^*}{\partial \beta} > 0 \quad \text{and} \quad \frac{\partial i_j^*}{\partial \beta} < 0$$

---

<sup>26</sup>By symmetry with case (iii), it is also possible to show that in equilibrium (viii) the informal employment rate is always larger for the majority group.

Appendix E proves these results. These findings confirm the intuition developed in the previous paragraph that less homophily or less segregation (*i.e.* a larger  $\beta$ ) increases the formal employment rate of each group. Once again, the underlying mechanism is that improving communication across groups enlarges individuals' network, as they gain access to the other group's information. Then information about formal job offers circulates more easily, raising formal employment prospects. More interestingly, these results reveal that reducing segregation actually lowers informal employment, although informal job offers also transit through word of mouth communication. This is because in this equilibrium, increasing  $\beta$  does not only enlarge the network, but also increases the probability that workers from group  $j$  (the group engaged in the informal sector) get access to formal job offers by interacting with members of group  $-j$ . Conversely, even if group  $-j$  members also become more exposed to informal job offers by interacting more with group  $j$  members, they are still in a situation where the incentives to decline such offers dominate.

This result is at odds with one of the findings of Calvó-Armengol et al. (2007) in a comparable setting. In their model, agents all belong to the same group and can either be employed, unemployed or criminals. Each individual is related to a unique other agent with whom s/he is *strongly* tied, and is otherwise weakly related to any other agent. At each period, agents are randomly matched by pair with their dyad partner or with any other agent, according to some given probability. One of the findings of the paper is that increasing the frequency of matching with weak ties increases the crime rate. This result stems from the fact that only crime opportunities circulate through the network: job offers arrive exogenously (e.g employment agencies) and are not transmitted through word of mouth. By contrast, I have shown that if the network plays a role in both sectors (employment and crime or formal and informal employment), then increasing social interactions can lower the participation in the "bad" sector provided that the alternative activity is attractive.

I finally look at how a policy aiming at directly fighting informal employment actually affects the various employment rates. In this model, such a policy corresponds to an increase in the audit rate  $p$ . Unsurprisingly, I can show that:

$$\frac{\partial f_{-j}^*}{\partial p} > 0, \quad \frac{\partial f_j^*}{\partial p} > 0 \text{ and } \frac{\partial i_j^*}{\partial p} < 0$$

However, the mechanisms behind these results differ from those explaining the effect of  $\beta$ . Variations in  $\beta$  alter the job information transmission process, and hence impact the various employment rates by influencing the probability of unemployed to hear about a job offer. On the other hand, variations in  $p$  directly affect

the incentives to engage in the informal economy in two ways. An increase in the apprehension rate both increases the rate at which informal jobs are terminated and decreases the net informal wage, affecting both the transition rates and the financial incentives. The effect of a rise in the audit rate on the formal employment rate is therefore indirect, and comes from the reduced incentives to informal work, while the effect of lower segregation directly improves the formal employment rate through facilitated information transmission. Hence, an important difference between an increase in  $p$  and an increase in  $\beta$  is that the former increases the unemployment rate of the targeted group (the one participating in the informal economy), while the latter does not. This implies that policies improving social cohesion and communication across groups could be more efficient and in any case less costly in terms of unemployment than direct policies aiming at destroying informal jobs. This is particularly relevant in this equilibrium where unemployment is the least remunerating activity.

## 6 Conclusion and discussion

The simple theoretical framework used in this paper helps rationalizing the idea of the informal economy as an economic safety net and the over-representation of minorities in the underground economy. The model predicts that even when the informal sector is less attractive than the formal one in terms of wages, equilibria where some agents engage in undeclared work may yet arise. More specifically, I show that all equilibria involving the informal sector arise when unemployment benefits are too low. In this case, some unemployed workers are better off accepting informal job offers than staying unemployed in order to wait for a formal job offer. In particular, in the equilibrium where only one group turns informal, this group is the minority group. The rationale for this directly lies in the process of job offers transmission and hence in the nature of social interactions. Because majority workers have access to a large network, they expect to receive formal job offers frequently enough to afford turning down informal job offers and staying unemployed a little longer. Conversely, minority workers receive too few formal offers to decline informal ones.

This model also proves insightful in terms of policy implications, would the planner's objective be to reduce the size of the informal sector. It suggests that increasing the level of the unemployment benefit (or more generally social welfare) could allow workers to remain unemployed and wait for a formal job offer rather than immediately accepting informal job offers. The interplay between formal employment, unemployment

and informal employment would thus deserve some attention when designing traditional employment policies. The model also shows that policies reducing homophily, that is enhancing inter-groups relationships and improving immigrants integration tend to favor formal over informal employment. More importantly, such policies turn out to be neutral in terms of unemployment, as they directly amplify the flow toward the formal sector. By contrast, policies directly combating undeclared jobs through more frequent audits seem less efficient: they do limit the size of the informal economy, but at the expense of a larger unemployment rate. Indeed, such policies boost the flow out of informality, but have no direct impact on the flow into formal employment.

Naturally, this model is not exempt from drawbacks. First, it does not feature a proper network in the sense of a graph with well-defined links between any two agents. Instead, in my model, any two agents can be connected or not at each period, and in this sense all agents are symmetric. Therefore, I am unable to draw conclusions regarding the place of an agent in the network (e.g. whether s/he is at the center or at the periphery of the graph), which has been proved quite relevant in the networks literature. Although this would undoubtedly refine the findings of the model, I am confident that the main results would be unaltered. Another concern may come from the fact that job offers, and in particular formal ones circulate only through the network. As already discussed, this assumption is highly unrealistic, and dropping it should significantly alter the dynamics of the model. In particular the exogenous rate of arrival formal job offers could be large enough to allow the minority workers to decline informal offers and remain unemployed until they receive a formal one. This of course could be mitigated if the model included the demand side of the labor market, which is obviously an important limitation of the model. Yet, even in this case, a possible and credible refinement would be to introduce discrimination, so that minority workers would receive fewer exogenous offers or be less often hired following a match. I am therefore convinced that most of the previous results would hold in this setting.

Finally, an interesting refinement would be to allow for different modes of communication across employment sectors. In particular, transmission of informal job offers could depend on trust. As pointed out by [Portes and Haller \(2005\)](#), trust is a key element in informal transactions: in as much as the informal economy is characterized by the lack of any written contract, trust is the only way to sustain informally established relationships and to overcome commitment problems. As it turns out, sharing a common identity,

and especially belonging to the same ethnic group is one of the best ways to generate such trust in informal exchanges.<sup>27</sup> Indeed, not only sanctions threats are more credible and hence more efficient within ethnic groups (Miguel and Gugerty, 2005), but migrants to a country also tend to be more socially cohesive: established contacts in the destination country is often a pre-requisite to immigration, and, once they arrive in a given country, migrants strengthen their links by developing community-based institutions (Munshi, 2003). Therefore, I could introduce a trust parameter that would allow informal job offers to circulate more easily among minority workers than among majority workers. This would most likely reinforce my results.

## A Characterization of the steady-states

Using equation (10) in equations (8) and (9), we get:

$$f_j^* = \frac{(f_j^* n_j + \beta f_{-j}^* n_{-j}) \alpha_F \phi_j}{\delta_F} (1 - f_j^* - i_j^*) \quad (18)$$

and

$$i_j^* = \frac{(i_j^* n_j + \beta i_{-j}^* n_{-j}) \alpha_I \psi_j}{\delta_I + p} (1 - f_j^* - i_j^*) \quad (19)$$

The values of  $\phi_j$  and  $\psi_j$  for which these two relationships are verified are derived below, for each steady-state.

(i) For  $f_j^* = 0$  and  $i_j^* = 0$ ,  $\forall j \in \{m, M\}$ , it is immediate to see that equations (18) and (19) are always verified for any value of  $\phi_j$  and  $\psi_j$  ( $\forall \phi_j, \psi_j \in \{0, 1\}$ ),  $\forall j \in \{m, M\}$ .

(ii) For  $f_j^* > 0$ ,  $f_{-j}^* = 0$  and  $i_j^* = i_{-j}^* = 0$ , equation (18) is verified for  $j$  if  $\phi_j = 1$ , and for  $-j$  if  $\phi_j = 0$ , and (19) is verified  $\forall \psi_j \in \{0, 1\}$ . Solving for equation (18) then leads to  $f_j^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{n_j} < 1$ . Note that we must have  $\frac{\delta_F}{\alpha_F} < n_j$  to ensure that  $f_j^* > 0$ .

(iii) For  $f_j^* > 0$ ,  $f_{-j}^* > 0$  and  $i_j^* = i_{-j}^* = 0$ , equation (19) is satisfied  $\forall \psi_j \in \{0, 1\}$ ,  $\forall j \in \{m, M\}$ , and we must

---

<sup>27</sup> Greif (1993) also established that the 11th century Maghribi traders organized into a coalition to overcome the commitment problem.

have  $\phi_j = \phi_{-j} = 1$  for equation (18) to hold. We can then rewrite equation (18) as follows,  $\forall j \in \{m, M\}$ :

$$f_j^* = \frac{(n_j f_j^* + \beta n_{-j} f_{-j}^*) \alpha_F}{\delta_F + (n_j f_j^* + \beta n_{-j} f_{-j}^*) \alpha_F} < 1 \quad (20)$$

In this equilibrium  $f_j^*$  and  $f_{-j}^*$  are particularly difficult to obtain analytically as they are determined by two equations of two unknowns (equation 20 for  $j$  and for  $-j$ ). Combining the two equations reveal that  $f_j^*$  actually solves  $\Phi(f_j^*) = 0$ , where  $\Phi(x)$  is a third order polynomial of the form:

$$\Phi(x) = A(1-x)^2 x + B(1-x)^2 + C(1-x)x + D(1-x) + Ex + F,$$

where  $A, B, C, D, E$  and  $F$  are constants made of the parameters  $n_j, n_{-j}, \beta, \alpha_F$  and  $\delta_F$ .

**(iv)** For  $f_j^* = 0, f_{-j}^* = 0, i_j^* > 0$  and  $i_{-j}^* = 0, \forall j \in \{m, M\}$ : as in case (i), equation (18) is verified  $\forall \phi_j \in \{0, 1\}, \forall j \in \{m, M\}$ . Then, equation (19) holds when  $\psi_j = 1$  and  $\psi_{-j} = 0$ . By symmetry with case (ii), we have  $i_j = 1 - \frac{\delta_{I+p}}{\alpha_I} \frac{1}{n_j}$ .

**(v)** For  $f_j^* > 0, f_{-j}^* = 0, i_j^* > 0$  and  $i_{-j}^* = 0, \forall j \in \{m, M\}$ , we must have  $\phi_j = 1, \phi_{-j} = 0, \psi_j = 1$  and  $\psi_{-j} = 0$  for equations (18) and (19) to hold. Re-arranging these equations leads to  $1 - f_j^* - i_j^* = \frac{\delta_F}{\alpha_F} \frac{1}{n_j}$  and  $1 - f_j^* - i_j^* = \frac{\delta_{I+p}}{\alpha_I} \frac{1}{n_j}$ , which in turn implies that  $\frac{\delta_F}{\alpha_F} = \frac{\delta_{I+p}}{\alpha_I}$ .

**(vi)** For  $f_j^* = 0, f_{-j}^* > 0, i_j^* > 0$  and  $i_{-j}^* = 0, \forall j \in \{m, M\}$ , we must have  $\phi_j = 0, \phi_{-j} = 1, \psi_j = 1$  and  $\psi_{-j} = 0$  for equations (18) and (19) to be satisfied. Then, we are back to case (iv) for group  $j$  with  $i_j = 1 - \frac{\delta_{I+p}}{\alpha_I} \frac{1}{n_j}$ , and to case (ii) for group  $-j$  with  $f_{-j}^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{n_{-j}}$ .

**(vii)** For  $f_j^* > 0, f_{-j}^* > 0, i_j^* > 0, i_{-j}^* = 0, \forall j \in \{m, M\}$ , equations (18) and (19) are verified with  $\phi_j = 1, \phi_{-j} = 1, \psi_j = 1$  and  $\psi_{-j} = 0$ . In this case, equation (19) for group  $j$  becomes:  $i_j^* = \frac{n_j i_j^* \alpha_I}{\delta_{I+p}} (1 - f_j^* - i_j^*)$ . Dividing both sides by  $i_j^*$  and re-arranging leads to  $i_j^* = 1 - f_j^* - \frac{\delta_{I+p}}{\alpha_I} \frac{1}{n_j}$ . We can then plug this expression of  $i_j^*$  in equation (18) for group  $j$  to obtain:  $f_j^* = \frac{(n_j f_j^* + \beta n_{-j} f_{-j}^*) \alpha_F}{\delta_F + (n_j f_j^* + \beta n_{-j} f_{-j}^*) \alpha_F} \frac{\delta_{I+p}}{\alpha_I} \frac{1}{n_j}$ . From this, we can express  $f_j^*$  as a function of  $f_{-j}^*$ :

$$f_j^* = \beta \frac{n_{-j}}{n_j} \frac{\alpha_F (\delta_{I+p})}{\alpha_I \delta_F - \alpha_F (\delta_{I+p})} f_{-j}^* \quad (21)$$

Note that because both sides must be positive, it imposes that  $\frac{\alpha_I}{\delta_I+p} > \frac{\alpha_F}{\delta_F}$ , *i.e.* that the net job arrival rate is larger in the informal sector. Finally, using the previous expression of  $f_j^*$  in equation (18) for group  $-j$  and re-arranging, allow me to derive the value of  $f_{-j}^*$  at this steady-state:

$$f_{-j}^* = 1 - \frac{\delta_F}{\alpha_F} \frac{\alpha_I \delta_F - \alpha_F (\delta_I + p)}{\alpha_I \delta_F - (1 - \beta^2) \alpha_F (\delta_I + p)} \frac{1}{n_{-j}}$$

Hence the value of  $f_j^*$ :

$$f_j^* = \beta \left[ n_{-j} \frac{\alpha_F (\delta_I + p)}{\alpha_I \delta_F - \alpha_F (\delta_I + p)} - \frac{\delta_F (\delta_I + p)}{\alpha_I \delta_F - (1 - \beta^2) \alpha_F (\delta_I + p)} \right] \frac{1}{n_j}$$

We can also deduce the value of  $i_j^*$  from the following relationship:  $i_j^* = 1 - \frac{\delta_I+p}{\alpha_I} \frac{1}{n_j} - f_j^*$ .

**(viii)** For the no formal sector steady state  $f_j^* = 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* > 0$ ,  $\forall j \in \{m, M\}$  to exist, equations (18) and (19) hold  $\forall \phi_j, \phi_{-j} \in \{0, 1\}$  and require that  $\psi_j = \psi_{-j} = 1$ ,  $\forall j$ . By symmetry with case (iii) we deduce the following relationship:  $i_j^* = \frac{(n_j i_j^* + \beta n_{-j} i_{-j}^*) \alpha_I}{\delta_I + p + (n_j i_j^* + \beta n_{-j} i_{-j}^*) \alpha_I}$ ,  $\forall j \in \{m, M\}$ .

**(ix)** For  $f_j^* > 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* > 0$   $\forall j \in \{m, M\}$ , equations (18) and (19) are satisfied if  $\phi_j = 1$ ,  $\phi_{-j} = 0$  and  $\psi_j = \psi_{-j} = 1$ ,  $\forall j$ . This case is symmetric to case (vii) in terms of activities. Re-arranging equation (18) for group  $j$  leads to  $f_j^* = 1 - i_j^* - \frac{\delta_F}{\alpha_F} \frac{1}{n_j}$ . Then, plugging this expression of  $f_j^*$  in equation (19) for group  $j$  allows us to express  $i_j^*$  as a function of  $i_{-j}^*$ :  $i_j^* = \beta \frac{n_{-j}}{n_j} \frac{\alpha_I \delta_F}{\alpha_F (\delta_I + p) - \alpha_I \delta_F} i_{-j}^*$ . Note that because both sides must be positive, it imposes that  $\frac{\alpha_I}{\delta_I+p} < \frac{\alpha_F}{\delta_F}$ , *i.e.* that the net job arrival rate is larger in the formal sector. Finally, using the previous expression of  $i_j^*$  in equation (19) for group  $-j$  and re-arranging, allow me to derive the value of  $i_{-j}^*$  at this steady-state:

$$i_{-j}^* = 1 - \frac{\delta_I + p}{\alpha_I} \frac{\alpha_F (\delta_I + p) - \alpha_I \delta_F}{\alpha_F (\delta_I + p) - (1 - \beta^2) \alpha_I \delta_F} \frac{1}{n_{-j}}$$

Hence the value of  $i_j^*$ :

$$i_j^* = \beta \left[ n_{-j} \frac{\alpha_I \delta_F}{\alpha_F (\delta_I + p) - \alpha_I \delta_F} - \frac{\delta_F (\delta_I + p)}{\alpha_F (\delta_I + p) - (1 - \beta^2) \alpha_I \delta_F} \right] \frac{1}{n_j}$$

We can also deduce the value of  $f_j^*$  from the following relationship:  $f_j^* = 1 - \frac{\delta_F}{\alpha_F} \frac{1}{n_j} - i_j^*$ .

(**x**) For the mixed steady-state  $f_j^* > 0$  and  $i_j^* > 0$ ,  $\forall j \in \{m, M\}$  to arise, equations (18) and (19) require that  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ . Then, we obtain the following relationships.

$$f_j^* = (n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{\delta_F} (1 - f_j^* - i_j^*), \quad \forall j \in \{B, W\}$$

and

$$i_j^* = (n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{\delta_I + p} (1 - f_j^* - i_j^*), \quad \forall j \in \{B, W\}$$

## B Equilibria

Remember that  $w_F(1 - t) > b$  by assumption. Using equations (16) and (17), I now provide the conditions for the different values of the choice variables to arise in equilibrium. In what follows, I denote  $w_F(1 - t)$  as  $w_F$  and  $w_I(1 - pf)$  as  $w_I$  to simplify the expressions.

(**i**) The full unemployment steady-state where  $f_j^* = 0$  and  $i_j^* = 0$ ,  $\forall j \in \{m, M\}$ , exists  $\forall \phi_j, \psi_j \in \{0, 1\}$ ,  $\forall j \in \{m, M\}$ . Given the assumption  $w_F > b$ , I cannot have  $\phi_j = 0$ , which requires that  $(r + \delta_I + p)[w_F - b] < 0$  from equation (16). So this steady-state cannot arise in equilibrium for  $\phi_j = 0$ ,  $\forall j \in \{m, M\}$ . Conversely, the condition for  $\phi_j = 1$  when  $i_j^* = i_{-j}^* = 0$  is given by:  $(r + \delta_I + p)[w_F - b] \geq 0$ , equivalent to  $w_F \geq b$ , which is always true. Note that the condition for  $\phi_{-j} = 1$  is exactly the same, hence I cannot have  $\phi_j \neq \phi_{-j}$ . So, we must have  $\phi_j = \phi_{-j} = 1$  for steady-state (i) to arise in equilibrium. I then consider two different cases according to the value of  $\psi_j$ :

- ◇ Case  $\psi_j = 1$  requires that  $(r + \delta_F)[w_I - b] \geq 0$  as  $f_j^* = f_{-j}^* = 0$ , equivalent to  $w_I \geq b$ . Note that we cannot have  $\psi_j \neq \psi_{-j}$  either. Combining what precedes, we can state that the full unemployment steady-state (i) arises with  $\phi_j = \phi_{-j} = \psi_j = \psi_{-j} = 1$ ,  $\forall j \in \{m, M\}$  if  $b \leq \min\{w_F, w_I\}$ .
- ◇ Case  $\psi_j = \psi_{-j} = 0$  requires  $w_I(1 - pf) < b$ . Combining what precedes, I can state that the full unemployment steady-state (i) arises with  $\phi_j = \phi_{-j} = 1$ , and  $\psi_j = \psi_{-j} = 0$  if  $w_I < b \leq w_F$ .

To summarize, the full-unemployment steady-state (i) always arises in equilibrium:

- ◇ when  $w_F > w_I$ , (i) arises in equilibrium:

- if  $b \leq w_I$ , with  $\phi_j = \phi_{-j} = \psi_j = \psi_{-j} = 1, \forall j \in \{m, M\}$ , or
- if  $w_I < b \leq w_F$ , with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 0, \forall j \in \{m, M\}$ .

◇ when  $w_F \leq w_I$ , (i) arises in equilibrium if  $b \leq w_F$ , with  $\phi_j = \phi_{-j} = \psi_j = \psi_{-j} = 1, \forall j \in \{m, M\}$ .

(ii) The steady-state where one group is fully unemployed while the other takes part in the formal labor market only ( $f_j^* > 0, f_{-j}^* = 0$  and  $i_j^* = i_{-j}^* = 0, \forall j \in \{m, M\}$ ) requires that  $\phi_j = 1, \phi_{-j} = 0$  (and  $\forall \psi_j \in \{0, 1\}$ ). Given that  $i_j^* = 0$ , the condition for  $\phi_{-j} = 0$  (equation 16) becomes  $(r + \delta_I + p)[w_F - b] < 0$ , which is impossible. Hence this steady-state never arises in equilibrium.

(iii) The no informal sector steady-state ( $f_j^* > 0, f_{-j}^* > 0$  and  $i_j^* = i_{-j}^* = 0$ ), exists for  $\phi_j = \phi_{-j} = 1$  and  $\forall \psi_j \in \{0, 1\}, \forall j \in \{m, M\}$ . The condition for  $\phi_j = \phi_{-j} = 1$  is always satisfied as  $i_j^* = i_{-j}^* = 0, \forall j \in \{m, M\}$ . We can then consider three different cases according to the values of  $\psi_j$  and  $\psi_{-j}$ :

◇ If  $\psi_j = \psi_{-j} = 1$ , the conditions from equation (16) are given by  $b \leq w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$ ,  $\forall j$ . Combining with the condition on  $\phi_j$  (i.e.  $b \leq w_F$ ), we get the following condition:

$$b \leq \min \left\{ \begin{array}{l} w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} \\ w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} \\ w_F \end{array} \right\}$$

It is easily shown that when  $w_F \leq w_I$ ,  $w_F$  is the minimum of the right-hand side term (henceforth RHS) in the above expression, while when  $w_F > w_I$ , the minimum is  $w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$ , where  $j$  and  $-j$  are such that  $n_j f_j^* > n_{-j} f_{-j}^*$ .

◇ If  $\psi_j = \psi_{-j} = 0$ , the various conditions become:

$$\max \left\{ \begin{array}{l} w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} \\ w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} \end{array} \right\} < b \leq w_F$$

We can deduce from the previous case that  $w_F > w_I$ . The maximum of the left-hand side term (henceforth LHS) is  $w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F}$ , where  $j$  and  $-j$  are such that  $n_j f_j^* > n_{-j} f_{-j}^*$ .

◇ If  $\psi_j = 0$  and  $\psi_{-j} = 1$ , all the conditions put together give:

$$w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} < b \leq \min \left\{ \begin{array}{l} w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} \\ w_F \end{array} \right\}$$

From the previous cases, we know that when  $w_F > w_I$ , then  $w_F$  is the maximum of the RHS. On the contrary, when  $w_F \leq w_I$ , we get a contradiction as  $w_F$  cannot be the minimum of the RHS and larger than the LHS at the same time.

To summarize, steady-state (iii) always arises in equilibrium:

◇ when  $w_F > w_I$ :

- if  $b \leq w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$ , with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ ,
- if  $w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} < b \leq w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F}$ , with  $\phi_j = \phi_{-j} = 1$ ,  $\psi_j = 0$  and  $\psi_{-j} = 1$ ,
- if  $w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} < b \leq w_F$ , with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 0$ ,

where  $j$  and  $-j$  are such that  $n_j f_j^* > n_{-j} f_{-j}^*$ .

◇ when  $w_F \leq w_I$ : if  $b \leq w_F$  (always true) with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ .

(iv) The steady-state where one group is fully unemployed while the other takes part in the informal labor market only ( $f_j^* = 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ ) exists for  $\psi_j = 1$ ,  $\psi_{-j} = 0$  and  $\forall \phi_j, \phi_{-j} \in \{0, 1\}$ . The condition (equation 17) for  $\psi_j = 1$  writes  $(r + \delta_F)(w_I - b) \geq 0$ , while that for  $\psi_{-j} = 0$  writes  $(r + \delta_F)(w_I - b) < 0$ , hence a contradiction. We can then conclude that this steady-state never arises in equilibrium.

(v) The steady-state where one group is fully unemployed while the other work on both formal and informal markets ( $f_j^* > 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* = 0$ ,  $\forall j \in \{m, M\}$ ) exists for  $\phi_j = 1$ ,  $\phi_{-j} = 0$ ,  $\psi_j = 1$  and  $\psi_{-j} = 0$ . Given that  $i_{-j}^* = 0$ , the condition for  $\phi_{-j} = 0$  (equation 16) cannot be satisfied under the assumption that  $w_F > b$ . So this steady-state never arises in equilibrium.

(vi) The steady-state where one group specializes in the formal economy while the other specializes in the informal economy ( $f_j^* = 0, f_{-j}^* > 0, i_j^* > 0$  and  $i_{-j}^* = 0, \forall j \in \{m, M\}$ ) exists for  $\phi_j = 0, \phi_{-j} = 1, \psi_j = 1$ , and  $\psi_{-j} = 0$ . The conditions for the previous four equalities to be satisfied can be summarized as follows:

$$\max \left\{ \begin{array}{l} w_F - (w_I - w_F)n_j i_j^* \frac{\alpha_I}{r + \delta_I + p} \\ w_F - (w_I - w_F)n_{-j} i_{-j}^* \frac{\alpha_I}{r + \delta_I + p} \end{array} \right\} < b \leq \min \left\{ \begin{array}{l} w_I \\ w_F \end{array} \right\}$$

It is easy to check that for any of the LHS term to be smaller than  $w_I$ , we must have  $w_F < w_I$ , while for any of the LHS term to be smaller than  $w_F$ , we must have  $w_F > w_I$ , hence a contradiction. Therefore, we can conclude that this steady-state never arises in equilibrium.

(vii) Steady-states where one group is involved in both sectors, while the other works in the formal sector only ( $f_j^* > 0, f_{-j}^* > 0, i_j^* > 0$ , and  $i_{-j}^* = 0, \forall j \in \{m, M\}$ ), exist for  $\phi_j = \phi_{-j} = \psi_j = 1$  and  $\psi_{-j} = 0$ . These four conditions can be summarized as follows.

$$w_I - (w_F - w_I)(n_{-j}f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} < b \leq \min \left\{ \begin{array}{l} w_F - (w_I - w_F)n_j i_j^* \frac{\alpha_I}{r + \delta_I + p} \\ w_F \\ w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} \end{array} \right\}$$

It is easily checked that for the LHS term to be smaller than any of the RHS terms, we must have  $w_F > w_I$ , and  $j$  and  $-j$  such that  $n_j f_j^* < n_{-j} f_{-j}^*$ . Then, under the previous condition that  $w_F > w_I$ , it is easy to show that  $w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$  is the minimum of the RHS.

To summarize, steady-state (vii) arises in equilibrium with  $\phi_j = \phi_{-j} = \psi_j = 1$  and  $\psi_{-j} = 0$ , when  $w_F > w_I$ , if

$$w_I - (w_F - w_I)(n_{-j}f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_F}{r + \delta_F} < b \leq w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$$

where  $j$  and  $-j$  are such that  $n_j f_j^* < n_{-j} f_{-j}^*$ .

(viii) The no formal sector steady state  $f_j^* = 0, f_{-j}^* = 0, i_j^* > 0$  and  $i_{-j}^* > 0, \forall j \in \{m, M\}$  exists  $\forall \phi_j, \phi_{-j} \in \{0, 1\}$  and for  $\psi_j = \psi_{-j} = 1$ . The condition for  $\psi_j = \psi_{-j} = 1$  implies that  $w_I \geq b$ . We can then consider three

different cases:

◇ If  $\phi_j = \phi_{-j} = 1$ , the conditions are given by  $b \leq \min \left\{ \begin{array}{l} w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_F - (w_I - w_F)(n_{-j} i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_I \\ w_F \end{array} \right\}$  It is easily shown that when  $w_F \leq w_I$ , the minimum of the RHS is  $w_F$ . Conversely, when  $w_F > w_I$ ,  $w_I$  is the minimum of the RHS.

◇ If  $\phi_j = \phi_{-j} = 0$ , the condition becomes

$$\max \left\{ \begin{array}{l} w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_F - (w_I - w_F)(n_{-j} i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} \end{array} \right\} < b \leq \min \left\{ \begin{array}{l} w_I \\ w_F \end{array} \right\}$$

Here we have a contradiction as the LHS terms cannot be smaller than  $w_I$  and  $w_F$  at the same time.

Hence this steady-state cannot arise if  $\phi_j = \phi_{-j} = 0$ .

◇ If  $\phi_j = 0$  and  $\phi_{-j} = 1$ , all the conditions put together give:

$$w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p} < b \leq \min \left\{ \begin{array}{l} w_F - (w_I - w_F)(n_{-j} i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_I \\ w_F \end{array} \right\}$$

From what precedes, we know that this leads to a contradiction. Hence this steady-state cannot arise if  $\phi_j = 0$  and  $\phi_{-j} = 1$ .

To summarize, steady-state (viii) arises in equilibrium:

◇ when  $w_F > w_I$ : if  $b \leq w_I$  with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ .

◇ when  $w_F \leq w_I$ : if  $b \leq w_F$  (always true), with  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ .

**(ix)** Steady-states where one group is involved in both sectors, while the other works in the informal sector only ( $f_j^* > 0$ ,  $f_{-j}^* = 0$ ,  $i_j^* > 0$  and  $i_{-j}^* > 0, \forall j \in \{m, M\}$ ), exist for  $\phi_j = 1$ ,  $\phi_{-j} = 0$ , and  $\psi_j = \psi_{-j} = 1, \forall j$ .

These four conditions, combined with the condition that  $b \leq w_F$  sum up to the following condition:

$$w_F - (w_I - w_F)(n_{-j}i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} < b \leq \min \left\{ \begin{array}{l} w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_I - (w_F - w_I) n_j f_j^* \frac{\alpha_F}{r + \delta_F} \\ w_I \\ w_F \end{array} \right\}$$

We can check that for the LHS term to be smaller than any of the RHS terms, we must have  $w_F \leq w_I$ , and  $j$  and  $-j$  such that  $n_j i_j^* < n_{-j} i_{-j}^*$ . Then, we can immediately deduce that  $w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p}$  is the minimum of the RHS.

To summarize, steady-state (ix) arises in equilibrium with  $\phi_j = 1$ ,  $\phi_{-j} = 0$ , and  $\psi_j = \psi_{-j} = 1$ , when  $w_F \leq w_I$ , if

$$w_F - (w_I - w_F)(n_{-j}i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} < b \leq w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p}$$

where  $j$  and  $-j$  are such that  $n_j i_j^* < n_{-j} i_{-j}^*$ .

(x) The mixed steady-state ( $f_j^* > 0$  and  $i_j^* > 0$ ,  $\forall j \in \{m, M\}$ ) exists for  $\phi_j = \phi_{-j} = 1$  and  $\psi_j = \psi_{-j} = 1$ . This leads to the following condition:

$$b \leq \min \left\{ \begin{array}{l} w_F - (w_I - w_F)(n_j i_j^* + \beta n_{-j} i_{-j}^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_F - (w_I - w_F)(n_{-j} i_{-j}^* + \beta n_j i_j^*) \frac{\alpha_I}{r + \delta_I + p} \\ w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F} \\ w_I - (w_F - w_I)(n_{-j} f_{-j}^* + \beta n_j f_j^*) \frac{\alpha_f}{r + \delta_F} \\ w_F \end{array} \right\}$$

The terms  $w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$ ,  $\forall j$  are the only terms that can be smaller to any other without getting a contradiction, when  $w_F > w_I$ . We can finally state that the minimum is given by  $w_I - (w_F - w_I)(n_j f_j^* + \beta n_{-j} f_{-j}^*) \frac{\alpha_F}{r + \delta_F}$ , with  $n_j f_j^* > n_{-j} f_{-j}^*$  and  $n_j i_j^* < n_{-j} i_{-j}^*$ .

## C The minority group turns informal

When  $w_F(1-t) > w_I(1-pf)$ , the group that engages in the informal sector in equilibrium (vii) is the minority group:  $f_j^* > 0$ ,  $i_j^* > 0$ ,  $f_{-j}^* > 0$  and  $i_{-j}^* = 0$  with  $n_{-j} > n_j$ .

*Proof.* Using the expression of  $f_{-j}^*$  (12) and the fact that  $f_{-j}^* > 0$ , we can easily show that

$$n_{-j} > \frac{\delta_F}{\alpha_F} \frac{\alpha_I \delta_F - \alpha_F(\delta_I + p)}{\alpha_I \delta_F - (1 - \beta^2)\alpha_F(\delta_I + p)} \quad (22)$$

Note that this inequality can also be obtained using expression (11) and  $f_j^* > 0$  for instance.

The proof then consists in showing that the RHS in (22) is larger than one half, implying that  $n_{-j} > \frac{1}{2}$  and hence that the group working in the formal sector only is the majority group.

Using expression (21) and the fact that  $n_{-j}f_{-j}^* > n_jf_j^*$ , we find that  $\frac{\delta_F}{\alpha_F} > (1+\beta)\frac{\delta_I+p}{\alpha_I}$ , which is equivalent to

$$2\frac{\delta_F}{\alpha_F} > 2 + 2\beta\frac{\delta_I+p}{\alpha_I} \quad (23)$$

We can then show that the following relationship holds:

$$2 + 2\beta\frac{\delta_I+p}{\alpha_I} > 1 + \beta^2\frac{\alpha_F(\delta_I+p)}{\alpha_I\delta_F - \alpha_F(\delta_I+p)} \quad (24)$$

Indeed, we know from (21) and  $n_{-j}f_{-j}^* > n_jf_j^*$  that  $\beta\frac{\alpha_F(\delta_I+p)}{\alpha_I\delta_F - \alpha_F(\delta_I+p)} < 1$ . This implies that  $\beta\frac{\alpha_F(\delta_I+p)}{\alpha_I\delta_F - \alpha_F(\delta_I+p)} - 2\frac{\delta_I+p}{\alpha_I} < 1 - 2\frac{\delta_I+p}{\alpha_I} < 1$ . Hence,  $\beta[\beta\frac{\alpha_F(\delta_I+p)}{\alpha_I\delta_F - \alpha_F(\delta_I+p)} - 2\frac{\delta_I+p}{\alpha_I}] < 1$  as  $0 < \beta < 1$  which is equivalent to  $2 + 2\beta\frac{\delta_I+p}{\alpha_I} > 1 + \beta^2\frac{\alpha_F(\delta_I+p)}{\alpha_I\delta_F - \alpha_F(\delta_I+p)}$ .

By transitivity, inequalities (23) and (24) finally imply that:

$$\begin{aligned} 2\frac{\delta_F}{\alpha_F} &> 1 + \beta^2\frac{\alpha_F(\delta_I+p)}{\alpha_I\delta_F - \alpha_F(\delta_I+p)} \\ \Leftrightarrow 2\frac{\delta_F}{\alpha_F} &> \frac{\alpha_I\delta_F - (1-\beta^2)\alpha_F(\delta_I+p)}{\alpha_I\delta_F - \alpha_F(\delta_I+p)} \\ \Leftrightarrow \frac{\delta_F}{\alpha_F} \frac{\alpha_I\delta_F - \alpha_F(\delta_I+p)}{\alpha_I\delta_F - (1-\beta^2)\alpha_F(\delta_I+p)} &> \frac{1}{2} \end{aligned} \quad (25)$$

## **D.4 The Strength and Use of Local Referral Networks: Evidence from France**

# The Strength and Use of Local Referral Networks: Evidence from France

Camille Hémet

Clément Malgouyres \*

Paris School of Economics

Paris School of Economics

March 2019

Preliminary and Incomplete

## Abstract

Using the French Labor Force Survey, which contains exhaustive information on close neighbors as well as employer administrative identifiers, we show that workers residing in the same block are more likely to work in the same firm than those living in nearby blocks, suggesting the presence of local referral networks. We provide several robustness checks against two main threats to identification: sorting and reverse causality. Finally, focusing on transitions, we show that referral networks are used most intensively for out-of-unemployment transitions rather than job-to-job transitions while most the existing work has focused on the latter.

---

\*Camille Hémet: [camille.hemet@psemail.eu](mailto:camille.hemet@psemail.eu), Clément Malgouyres, [cmalgouyres@gmail.com](mailto:cmalgouyres@gmail.com). This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the “Investissements d’avenir” program (reference : ANR-10-EQPX-17 – Centre d’accès sécurisé aux données – CASD)

# 1 Introduction

A large body of empirical research shows that informal networks play a pervasive role in modern labor markets. Early studies based on survey data typically find that most unemployed resort to informal methods to search for jobs and that a large share of employees states having found their jobs through such methods. For instance, the seminal study by [Holzer \(1988\)](#) shows, based on a sample of unemployed youth in the United-States, that 90 % of them are searching for a job through a friend or relative while 50 % only rely on state agency.<sup>1</sup> A more recent strand of the literature relies on administrative data and matched employer-employee datasets to indirectly detect informal referral networks. The seminal work by [Bayer \*et al.\* \(2008\)](#) shows that direct neighbors, i.e. people living in the same census block, are much more likely to work in the same census block than residents of nearby blocks. The tendency of close neighbors to display an excess clustering in the same workplace on the US labor market is confirmed by [Hellerstein \*et al.\* \(2011\)](#), who use firm rather than work-location to determine coworkers. This evidence suggests that individuals share employment opportunities with their immediate neighbors and that overall, referral networks have an important spatial dimension. Yet, the literature does not provide a clear picture regarding the initial employment status of individuals finding a job through the network. More precisely, whether employed workers act as referrals for employed or unemployed acquaintances remains an open question. Similarly, we lack information on whether employed workers rely more heavily on and benefit more from referral networks than unemployed individuals.

In this paper, we present new evidence on the role of neighborhood-based referral networks in the case of France. Using specific features of the Labor Force Survey in France, we are able to use the approach developed in [Bayer \*et al.\* \(2008\)](#) to assess the strength of local referral network. Based on a very fine definition of housing block (20 to 40 households), we measure the strength of informal hiring network as the differential propensity of individuals living in the same block to work in the same firm in comparison with residents of a contiguous block, living so to speak “across the street”.<sup>2</sup> In particular, we contribute to the existing literature on local referrals by measuring how this strength differs across

---

<sup>1</sup>Job search methods of are not mutually exclusive. The author distinguished five different search methods: friends and relatives, responses to newspaper advertisements, state employment agencies, direct applications to employers, and “other” methods. These categories are those commonly used in labor force surveys, see e.g. [Addison and Portugal \(2002\)](#).

<sup>2</sup>More details on the definition of neighborhoods will be given in Section 2.

employment statuses. More specifically, after providing evidence on the existence of informal hiring networks at the very local block level, we use the short panel dimension of our data to show that this referral effect is particularly important for workers who were initially unemployed. By contrast, the effect turns out to be insignificant when we focus on job-to-job transitions. Although this result is still preliminary, it could suggest that direct neighbors are particularly useful for individuals who are out of employment since they can convey information that employers cannot infer from the candidates' employment status.

This exercise is threatened by endogenous residential sorting. Indeed, we do observe that individuals that are similar in terms of education or nationality for instance tend to live in the same group of contiguous blocks, referred to as a *sector*. Our identifying assumption consist in ruling out the possibility of sorting *within-sector*, while allowing for the alleged sorting *across sectors*. We carry out a test of this hypothesis in the spirit of [Altonji et al. \(2005\)](#), that supports its credibility. In turn, this assumption implies that after including sector fixed-effects, there is no block-level correlation in unobserved attributes among block residents, so that we are left to identify proper neighborhood effects.

We find that being block-neighbors increases the probability of being coworkers relatively to neighbors living in nearby block by 0.35 % on average, which represents about a third of the baseline probability, a figure remarkably close to available evidence for the US ([Bayer et al., 2008](#)). This is particularly striking as the French *sectors* (our reference group of contiguous blocks) is even more narrow than the block groups used in their study. As a matter of fact, most of the evidence on the existence of neighborhood-based referral network is based on American data. Evidence for Continental Europe and particularly France is much more scarce and we incidentally contribute to fill this gap.<sup>3</sup>

We are also able to provide some evidence that our results are not driven by reverse causality, i.e. the possibility that coworkers might referring each other on the housing market. To this aim, we restrict our estimation sample to public housing tenants. Given the shortage of public housing as well as the specific allocation mechanism of public housing units in France ([Algan et al., 2016](#)), reverse causality is highly unlikely for this specific subset of the population. The result presented above remains unchanged when estimated on the population of public housing tenants.

---

<sup>3</sup>An exception is [Hawranek and Schanne \(2015\)](#) who study residential job referral networks in Germany using matched employer-employee data but do not look at job transitions.

Our paper relates to a large literature on job search networks. As pointed out by [Ioannides and Datcher Loury \(2004\)](#) in their excellent survey, most of the existing evidence on such networks comes from surveys where workers are asked how they found their current job. Alternatively, more recent research such as [Bayer \*et al.\* \(2008\)](#) or [Hellerstein \*et al.\* \(2011\)](#) does not use direct information on the job search method used, but instead provides indirect evidence on the existence of job search methods by revealing excess clustering of coworkers in residential locations. In the same spirit, [Kramarz and Skans \(2014\)](#) find that firms are more likely to hire children of current employees than otherwise comparable job candidates, and [Oyer and Schaefer \(2016\)](#) present evidence that partners hire graduates from their own law school with a much higher probability than randomization would predict. In a recent paper, [Dustmann \*et al.\* \(2016\)](#) rely on very rich data combining matched employer-employee data with survey data and are therefore able to provide more direct evidence of referral-based job search networks, where networks are defined along ethnic lines. In particular, they show that workers earn higher wages and are less inclined to leave their firms if they have obtained their job through a referral. However, none of this paper is able to investigate empirically the heterogeneity of job referral networks across initial occupations.<sup>4</sup>

The remainder of the paper is organized as follows. Section 2 is devoted to presenting our data. In particular, we describe the specificities of the French Labour Force Survey that enable us to perform our analysis. Our empirical design is presented in Section 3, where we also provide a test of our identification assumption. Section 4 eventually exposes the baseline results, as well as the results based on job transitions. Section 5 concludes.

## 2 Data and institutional setting

### 2.1 The French Labor Force Survey and its clustered sampling scheme

In this paper, we use the French Labor Force Survey (*Enquête Emploi*, hereafter the LFS), which is conducted quarterly by the French statistical institute (INSEE) since 2003. One sixth of the sample is renewed each quarter, so that the survey takes the form of a rotating-panel, as each household is surveyed for six consecutive waves before leaving the sample. Each wave of the survey comprises about 72,000

---

<sup>4</sup>Although [Dustmann \*et al.\* \(2016\)](#) consider job-to-job and unemployment-to-job flows in their theoretical model, this is not the focus of their empirical analysis

respondents of age 15 or more. The LFS contains precise information about individuals' labor market situation: employment status, wage, type of contract, tenure, job search methods and socio-economic category. It also provides relevant individual characteristics, such as age, gender, education, and marital status. Importantly for our analysis, the LFS also contains the firm identifier for each employee, which enables us to identify whether two individuals actually work in the same firm. This information is similar to that used by [Hellerstein et al. \(2011\)](#) and constitutes an improvement over ? where co-workers are defined as individuals who work in the same block, a proxy for actual co-workers.

The sampling strategy of the LFS makes it particularly valuable for studying neighborhood effects. Mainland France is initially divided into primary units, which are themselves split into *sectors*. A sector is defined so as to contain between 120 and 240 homes, and its boundaries are following urban boundaries such as roads and railways. One sector is selected in each primary unit, with a probability proportional to the number of homes it comprises. Each selected sector is then divided into neighborhoods of twenty households on average, that we will refer to as *blocks* in the remainder of the paper. Six blocks are then randomly picked within each sector, when the sector comprises more than six blocks. Otherwise, all six blocks are included in the baseline sample. The rotating panel survey is then organized in the following way: in each sector, one of the six blocks is *exhaustively* surveyed over six quarters, at which point a second block of the same sector enters the sample for the six subsequent quarters. This process goes on until all six neighborhoods of a given sector are surveyed.<sup>5</sup>

This peculiar sampling strategy, designed to save money and time, has two important implications for our study. First, because the sample consists in a random sample of blocks in which all inhabitants are exhaustively surveyed, we can perfectly observe all individuals living in these neighborhoods and hence characterize the socio-economic and demographic environment of each surveyed individual. In particular, we know the exact labor market situation of each block neighbor during the six quarters over which the block is surveyed. Second, because six blocks are surveyed in each sector, it implies that we can use the variation of individuals' characteristics across blocks and *within-sector* in order to limit issues related to sorting.

---

<sup>5</sup>For more details on the sample composition and selection, please refer to [INSEE documentation](#)

However, an important issue given our purpose is that sectors are sampled one block at a time. This means that individuals belonging to the *same sector* and *different blocks* are never interviewed during the same quarter. To cope with this issue with minimum measurement error, we consider only the first and last (sixth) wave of interrogation of each block, and define a *subsector* as a two distinct blocks belonging to the same sector and interviewed one quarter apart. To be clear, a subsector is thus a the concatenation of individuals surveyed during the sixth wave of a given block and individuals surveyed in the first wave of the next block in the list. It means that each block is included in two subsectors: one with the last wave of the previous block (combined with the reference block's first wave), and one with the first wave of the subsequent block (combined with the reference block's last wave). For a more thorough sample description and subsector definition, we refer the reader to our previous work [Hémet and Malgouyres \(2018\)](#) where we use the same data and a similar strategy to identify the effect of neighborhood ethnic diversity on individuals' employment probability.

The way we define subsectors results in rather small reference neighborhoods, compared to initial sectors. This might limit our ability to detect small effects but ensures that the subcomponents of each subsector, *i.e.* the blocks, are very similar. Moreover, focusing on blocks sampled one quarter apart limits the problems caused by firm mortality over time which would lead us to over-estimate the impact of being block-neighbors on working in the same firm.<sup>6</sup>

## 2.2 Sample description

Because we are interested in block neighbors' relative propensity to work in the same firm, we only keep working age individuals (between the age of 25 and 59 years old) who are employed in a firm with a correct employer ID at the time of the survey. This leaves us with a baseline sample of 116,859 workers surveyed between the first quarter of 2003 and the last quarter of 2012.

In order to study the propensity of close neighbors to work together, we follow ? and examine the propensity of a *pair* of individuals to work in the same firm. More precisely, we compare this propensity for pairs of individuals that reside on the same block versus nearby block. Therefore, our unit of obser-

---

<sup>6</sup>Every year a substantial share of firms in France go through changes in their official ID-number. Such changes do not always imply that these firms are actually closing or opening. In fact, a majority of changes in ID-number are related to firm demographic operations which are associated with very little discontinuities in the labor market situation of its employees [Picart \(2008\)](#).

vation is a pair of individuals. More precisely, we build a sample of pairs of neighbors living in the same sector: each resident of a given sector is matched to every single other sector-neighbor included in the sample. This imply that individuals are matched both to direct block-neighbors, but also to neighbors from the contiguous block included in the sector. We exclude pairs of members from the same household, who are not considered as neighbors, which could lead us to over-estimate neighborhood effects. As explained above in Section 2.1, we focus on the first and last interrogations of each individual, so that they can be matched to neighbors from another block belonging to the same sector with a gap of only one quarter in the period of observation. Otherwise stated, we restrict our sample to pair of individuals belonging to the same *subsector*.

Our final sample consists in about 3,2 millions pairs of subsector-neighbors, who live in about 10,000 different blocks and about 5,000 subsectors. Blocks and sectors contain an average of 15 and 30 employed individuals respectively. Table 1 displays key summary statistics regarding the sample of individuals and pairs.

### 3 Empirical Analysis

#### 3.1 Estimating the strength of local referrals

We start by estimating the following specification:

$$W_{ij} = \alpha R_{ij}^a + \rho_{s(ij)} + \varepsilon_{ij} \quad (1)$$

where  $W_{ij}$  and  $R_{ij}^a$  are binary variables equal to one if, respectively, individuals  $i$  and  $j$  work in the same firm and live in the same block  $a(ij)$ . The term  $\rho_{s(ij)}$  is a subsector-fixed effect associated with the subsector  $s(ij)$  that includes the block  $a(ij)$ .

In equation 1, the coefficient  $\alpha$  captures the extent to which block-neighbors tend to work in the same firm relatively to subsector-neighbors. We take the propensity to work in the same location as an indication that one member of the pair provided a referral (or more generally information) to the other member about jobs available in her place of work, so that  $\alpha$  is our (indirect) measure of the strength of local referrals. Given the small size of subsectors (two blocks of 20 households on average) and the geographical contiguity of blocks within a sector, one can safely rule out that differences in the propensity

to work in the same firm arises because of different access to employment or transport infrastructure, factor which have been shown to be important determinants of labor market outcomes in urban settings.<sup>7</sup> Therefore the crux of the approach is to neutralize most obvious determinants of neighbors clustering within the same firm by using subsector-level fixed-effect, and interpret the remaining *within-subsector* differences between blocks in propensity to work in the same firm as evidence of informal hiring network.

### 3.2 Testing for the absence of sorting within subsectors

The strategy proposed to identify neighborhood effects hinges on two main assumptions. The first is that local interactions between neighbors that shape referral-networks are more local than other factors affecting the likelihood of two individuals to work in the same firm. A violation of that assumption would lead us to find no effect. The second assumption is that there is no within-subsector sorting on unobservable characteristics that could be driving the probability of two block-neighbors to work in the same firm. A violation of this assumption would lead us to overestimate the importance of local referral network. While this assumption is intrinsically unverifiable, this section proposes a test based on observed covariates. Indeed, the presence of sorting on observables would make the assumption of no sorting on unobservables particularly untenable while its absence would strengthen its plausibility.

To this aim, we rewrite the baseline model in 1 in terms of deviations from sector-average:

$$\widetilde{W}_{ij} = \alpha \widetilde{R}_{ij}^a + \widetilde{\varepsilon}_{ij} \quad (2)$$

where  $\widetilde{z}$  refers to the residual of from a regression of  $z$  on a set of sector-dummies.

The identifying assumption, which is not directly testable, writes as:

$$Corr(\widetilde{\varepsilon}_{ij}, \widetilde{R}_{ij}^a) = 0 \quad (3)$$

To support this assumption, in the spirit of [Schmutte \(2015\)](#), we attempt to test to what extent individuals seem to be sorting on observable characteristics. In order to perform this test, we start by randomly

---

<sup>7</sup>See (e.g. [Gobillon et al., 2011](#)) for an analysis of spatial mismatch in the Parisian region.

selecting one individual per block. We then regress a given observable characteristic  $x_i$ , say years of education, on the block average of that same variable  $x_{a(i)}$ , excluding the selected individual in the computation of the block-average. At this stage, we expect to observe an important degree of selection. We then run the same regression, except that we now include subsector fixed-effect, i.e. we regress  $\tilde{x}_i$  on  $\tilde{x}_{a(i)}$ .

A large degree of sorting across sectors, i.e. large value of  $R^2(x_i, x_{a(i)})$ , seems plausible and is not incompatible with our identification strategy. However we expect a very low degree of within-subsector sorting, that is to say low values for  $R^2(\tilde{x}_i, \tilde{x}_{a(i)})$ .

We repeat this procedure 100 times and compute the average R-squares for the regressions without and with subsector fixed-effects. The corresponding results for several variables of interest are displayed in columns 1 and 2 of Table 5 respectively. We see clearly that college-educated individuals tend to live in blocks with a high share of college educated neighbors: the average R-square over the 100 repetitions is of 6.83 %. However, including sector fixed effects divides the R-square by a factor of about 10. It seems therefore that the amount of within-subsector sorting based on college-education is negligible. The same holds for other levels of education and a dummy variable for foreign nationality. A similar test is performed in [Hémet and Malgouyres \(2018\)](#) who use the same data and rely on the same identifying assumption. Although their sample is different as it includes all individuals and not pairs of employed workers, their results also support the assumption of no sorting on observables across blocks within subsectors.

## 4 Results

### 4.1 Baseline results on residential referral networks

Table 3 present the baseline results. Column (1) shows that being block neighbors is associated with a 0.35% increase in the probability of being coworker. Given an unconditional probability of about 1, the effect represents roughly a third of the baseline probability, a magnitude very similar to what has been found on US data by [Bayer et al. \(2008\)](#).

Column (2) focuses only on workers who have been employed in their firms for a quarter or more,

in order to solve issues related to systematic measurement error due to inter-quarter firm mortality which would tend to bias upward our estimator of the referral network effect.

Columns (3) to (5) use information on tenure and/or occupational status to restrict the sample to pairs  $(i, j)$  such that  $i$  has a shorter tenure (Column 3) than  $j$ , a lower occupational status  $j$  (Column 4), or both (Column 5). The coefficients are very stable over the three columns.

A potential issue in the previous results is that although we do observe spatial clustering of co-workers (as in Hellerstein *et al.*, 2011), there is a risk of reverse causality. More precisely, we cannot completely rule out that individuals initially working in the same firm but living in distinct neighborhoods shared information about residential opportunities and ended up living in the same neighborhood. In order to address this issue, we alternatively focus on a subsample of public housing tenants. Indeed, given the highly centralized allocation mechanism of public housing units and the relative shortage of such housing, it is highly unlikely that individuals working for the same employer will meet in the work place and then refer each other on the public housing market.<sup>8</sup> The results obtained on this specific population are displayed in Column (6). Overall, we find larger referral effect in this subsample. However, once this effect is expressed as a percentage of the baseline probability, its magnitude is virtually identical to that of previous columns. An alternative way of bypassing this potential reverse causality issue is proposed in Section 6 below.

Finally, we look at the heterogeneity of this relationship along the economic cycle. More precisely, we split the sample between the period prior the Great Recession and during this economic downturn, where the cutoff date is set at the first quarter of 2009. The corresponding results are displayed in Table 4. They suggest that the strength of local social interactions is lower in the post-2009 period.

## 4.2 Job transitions

We now turn to identifying local referral networks on job transitions, rather than on the stock of all employed workers. More precisely, we take advantage of the short longitudinal dimension of our data set to run a second set of estimations based on the sample of individuals that we observe entering a new firm

---

<sup>8</sup>See Algan *et al.* (2016) for a more in depth discussion of assignment to public housing units.

between two survey waves. These individuals are thus considered taking up a new job between two quarters. Looking at transitions therefore comes at the expense of sample size, but is valuable for two reasons. First, it allows us to circumvent the issue of reverse causality: if we observe an individual starting to work in one of her neighbors' firm, it is not possible that they ended up living in the same neighborhood after being co-workers. In addition, it enables us to shed some light on the type of individuals (unemployed or employed) that is the more likely to benefit from very local networks of referrals.

As argued in the introduction, this question is absent from the empirical literature. Theoretical predictions regarding whether referral network use is more prevalent for out-of-non employment versus job-to-job transitions is also ambiguous. Referral network are often considered as a way to reduce the noise around the signal relating to the productivity of an applicant (Dustmann *et al.*, 2016; Brown *et al.*, 2016). Being employed itself conveys information about a worker's productivity and implies that she has access to a professional network. According to this line of argument, one would expect residential referral network to be more intensively used by individuals who are currently not employed. However, being unemployed could also be associated with a stigma among neighbors thus causing referrals to be harder to obtain. The data at hand allows us determine whether residential referral network are more intensively used for job-to-job or non employment-to-job transitions. To this aim, we restrict the sample on individuals who were either active prior to finding their current job.

The baseline equation is identical to equation 1, except that the dependent variable  $TW_{ij}$  now is equal to one when individual  $i$  finds a new job in the same company as  $j$ .

$$TW_{ij} = \alpha R_{ij}^a + \rho_{s(ij)} + \varepsilon_{ij} \quad (4)$$

In order to preserve a sample of size that is large enough, we consider all labor market transitions even if they do not correspond to the first or last quarter of observation of the block. As a consequence, we slightly need to adapt the identification strategy proposed in section 3 to this new sample. In this section, we will define the subsector, *i.e.* the reference neighborhood where households chose to reside, as the set of (two) blocks that are surveyed at most three quarters apart. That is, we consider as individual  $i$ 's subsector-neighbors all individuals  $j$  living in the same sector and surveyed within a gap of 3 quarters or less with respect to  $i$ 's transitions.

We first estimate this equation on the overall sample of all job transitions, to check whether the baseline effect measured in section 4.1 are robust to this alternative sample. We then alternatively estimate  $\alpha$  on the subsample of job-to-job transitions and on the subsample of unemployment-to-job transitions. The corresponding results are displayed in Table 6. The baseline estimate in column (1) shows that local referral effects are still detectable among job transitions. Importantly, the magnitude of the effect is very comparable to that obtained in the first column of Table 3: here again, the estimated effect represents about a third (28.5 %) of the baseline probability to work in the same firm for subsector-neighbors. The results displayed in columns (2) and (3) are estimated on job-to-job transitions and on unemployment-to-job transitions respectively. They reveal that local referral network seems to be more relevant for the latter type of transitions rather than for the former. More precisely, the local referral effect estimated on the sample of unemployment-to-job transitions represents 60 % of the baseline probability.

Although these results are still preliminary, they could suggest that direct neighbors are particularly useful for individuals who are out of employment since they can convey information that employers cannot infer from the candidates' employment status. This result can also be helpful to understand why local referrals seem less relevant in times of economic turmoil, as documented at the end of Section 4.1. Indeed, if these networks are particularly relevant for unemployed individuals, this might result in some congestion during economic crisis where the number of unemployed neighbors increases and the number of job offers declines, explaining why they become less intensively relied upon. Numerous studies based on administrative data focus on job-to-job transitions – sometimes due the poor quality of data for people outside of employment (e.g. Schmutte, 2015). Our finding suggests they might be missing an important role of local labor market referrals.

## 5 Conclusion

In this paper, we find evidence of local referral network effects in France that are of comparable magnitude to what has been found in the case of the US (Bayer *et al.*, 2008). We show that this effect is robust to numerous checks regarding the issue of sorting and reverse causality.

The presence of local referrals is a possible way to explain the presence of well-documented neighborhood effects. For instance, [Solignac and To \(2014\)](#) focus on young graduates in France and find that local (neighborhood) employment rate affects positively job access. Such positive neighborhood effect in terms of labor market entry might be partially driven by the direct referrals we detect here.

Focusing on labor market transitions, we are still able to capture these referral network effects. More importantly, we find that such effects are concentrated on unemployment-to-job transitions and are not significant for job-to-job transitions. To our knowledge, we are the first to investigate the strength of local referral networks across initial employment status. If this result is confirmed by future robustness analysis, it can provide interesting insights regarding the way such local information networks are used.

## 6 Tables

Table 1: Descriptive statistics on individuals and pairs

Variable	mean
Individuals	
No degree	29.45
Prof. Degree	33.24
High-school	16.52
College	20.80
Foreign	3.88
Male	55.37
25-34 year old	27.70
35-44 year old	31.63
45-59 year old	40.67
# obs	116,859
Pairs	
Coworker	1.06
Block Neighbors	49.63
# obs	3,262,782
# <i>blocks</i>	10,080
# <i>sectors</i>	5,040

Notes: .

Table 2: Sorting on observables, with and without “sector” fixed-effect:  $R^2$  of bivariate regressions (in %)

Variable	$R^2(x, x_a)$	$R^2(\tilde{x}, \tilde{x}_a)$
No degree	2.74	0.53
High-school	1.91	0.31
College	6.83	0.72
Foreign	2.22	0.67

Notes: We randomly select one individual per block. For the set of displayed variables, we regress the variable on the block average (leaving the selected individual out) without and with subsector level fixed effect. We repeat this procedure 100 times and stored the average R-squares, without (Column 1) and with fixed-effect (Column 2).

Table 3: Baseline results on direct referrals

	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Baseline</i>	<i>Ten. &gt; 1Q</i>	<i>LowTen.</i>	<i>LowOcc.</i>	<i>LowOcc.Ten.</i>	<i>Pub.Hsg</i>
$R$	0.355*** (8.589)	0.364*** (8.51)	0.335*** (8.34)	0.388*** (8.89)	0.375*** (8.19)	0.493*** (4.80)
$\bar{W}$	1.061	1.082	1.035	1.145	1.104	1.36
Obs.	3262782	3170135	1601471	2552779	1263868	332926

Notes: Obs. refers to the number of pairs of neighbors used in the estimation. Standard errors are cluster at the sector-level. Column (1) presents results based on the entire sample. (2) restrict the sample to individuals who have a firm-tenure longer than 1 quarter. (3) restricts the sample to pairs such that  $i$  has a lower job tenure than  $j$ . (4) restricts the sample to pairs such that  $i$  has a hierarchically-lower occupation than  $j$ . (5) Subsample that fits the intersection of the conditions in 3 and 4. (6) restricts the sample to public housing tenants.

Table 4: Results on direct referrals: before and during the Great Recession

Prior	(1)	(2)	(3)	(4)	(5)	(6)
	<i>Baseline</i>	<i>Ten. &gt; 1Q</i>	<i>LowTen.</i>	<i>LowOcc.</i>	<i>LowOcc.Ten.</i>	<i>Pub.Hsg</i>
$R$	0.431*** (7.01)	0.441*** (6.93)	0.407*** (6.84)	0.471*** (7.28)	0.452*** (6.61)	0.490*** (4.22)
$\bar{W}$	1.179	1.203	1.149	1.274	1.149	1.389
Obs.	2043211	1985444	1002988	1599986	792634	227671
During	<i>Baseline</i>	<i>Ten. &gt; 1Q</i>	<i>LowTen.</i>	<i>LowOcc.</i>	<i>LowOcc.Ten.</i>	<i>Pub.Hsg</i>
$R$	0.229*** (5.85)	0.236*** (5.83)	0.215*** (4.53)	0.252*** (4.97)	0.245*** (4.89)	0.499** (2.47)
$\bar{W}$	0.863	0.88	0.844	0.93	0.844	1.298
Obs.	1219571	1184691	598483	952793	471234	105255

Notes: Obs. refers to the number of pairs of neighbors used in the estimation. Standard errors are cluster at the sector-level. The onset of the Great Recession is timed at 2009Q1. Column (1) presents results based on the entire sample. (2) restrict the sample to individuals who have a firm-tenure longer than 1 quarter. (3) restricts the sample to pairs such that  $i$  has a lower job tenure than  $j$ . (4) restricts the sample to pairs such that  $i$  has a hierarchically-lower occupation than  $j$ . (5) Subsample that fits the intersection of the conditions in 3 and 4. (6) restricts the sample to public housing tenants.

Table 5: Results on local referral effects among job-transitions

	(1)	(2)	(3)
	<i>Baseline</i>	J-to-J	U-to-J
<i>R</i>	0.123** (2.09)	0.0411 (0.61)	0.342*** (2.67)
$\bar{W}$	0.432	0.373	0.563
Obs.	88968	61621	27347

Notes: Results based on transitions. Block-neighbors are observed contemporaneously and sector-neighbors are observed with a gap of at most 3 quarters with respect to quarter  $t$  when the transition occurs between  $t$  and  $t + 1$ .

## References

- ADDISON, J. T. and PORTUGAL, P. (2002). Job search methods and outcomes. *Oxford Economic Papers*, **54** (3), 505–533.
- ALGAN, Y., HÉMET, C. and LAITIN, D. D. (2016). The social effects of ethnic diversity at the local level: A natural experiment with exogenous residential allocation. *Journal of Political Economy*, **124** (3), 696–733.
- ALTONJI, J. G., ELDER, T. E. and TABER, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, **113** (1), pp. 151–184.
- BAYER, P., ROSS, S. L. and TOPA, G. (2008). Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy*, **116** (6), 1150–1196.
- BROWN, M., SETREN, E. and TOPA, G. (2016). Do informal referrals lead to better matches? evidence from a firm’s employee referral system. *Journal of Labor Economics*, **34** (1), 161–209.
- DUSTMANN, C., GLITZ, A., SCHÖNBERG, U. and BRÜCKER, H. (2016). Referral-based Job Search Networks. *The Review of Economic Studies*, **83** (2), 514–546.
- GOBILLON, L., MAGNAC, T. and SELOD, H. (2011). The effect of location on finding a job in the paris region. *Journal of Applied Econometrics*, **26** (7), 1079–1112.
- HAWRANEK, F. and SCHANNE, N. (2015). *Your very private job agency: Job referrals based on residential location networks*. University of Regensburg Working Papers in Business, Economics and Management Information Systems 483, University of Regensburg, Department of Economics.
- HELLERSTEIN, J. K., MCINERNEY, M. and NEUMARK, D. (2011). Neighbors and Coworkers: The Importance of Residential Labor Market Networks. *Journal of Labor Economics*, **29** (4), 659 – 695.
- HÉMET, C. and MALGOUYRES, C. (2018). Diversity and Employment Prospects: Neighbors Matter! *Journal of Human Resources*, **53** (3), 825–858.

- HOLZER, H. J. (1988). Search Method Use by Unemployed Youth. *Journal of Labor Economics*, **6** (1), 1–20.
- IOANNIDES, Y. M. and DATCHEL LOURY, L. (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, **42** (4), 1056–1093.
- KRAMARZ, F. and SKANS, O. N. (2014). When Strong Ties are Strong: Networks and Youth Labour Market Entry. *Review of Economic Studies*, **81** (3), 1164–1200.
- OYER, P. and SCHAEFER, S. (2016). Firm/employee matching: An industry study of u.s. lawyers. *ILR Review*, **69** (2), 378–404.
- PICART, C. (2008). Flux d'emploi et de main-d'œuvre en France : un réexamen. *Économie et Statistique*, **412** (1), 27–56.
- SCHMUTTE, I. M. (2015). Job Referral Networks and the Determination of Earnings in Local Labor Markets. *Journal of Labor Economics*, **33** (1), 1 – 32.
- SOLIGNAC, M. and TO, M. (2014). *Neighborhood Effect and Labor Market Integration*. ERSA conference papers ersa14p1734, European Regional Science Association.

## **D.5 Unemployment and Crime Victimization: A Local Approach**

# Unemployment and Crime Victimization: a Local Approach

CAMILLE HÉMET\*

November 2018

## Abstract

This paper explores the relationship between unemployment rate and crime victimization at the neighborhood level, using data from the French victimization survey. The very local nature of the data enables me to tackle the endogenous location selection issue: once I control for the characteristics of a larger area into which household select their location, the remaining variation of observables across neighborhoods within this larger area can be considered as exogenous. The contribution of this paper to the economics of crime literature is then twofold. First, I show that, at the very local neighborhood level, unemployment rate is an important factor explaining victimization. Second, I take advantage of the precise localization of the data to compare the effect of unemployment rate in the reference neighborhood and in adjacent neighborhoods. The results support the idea that criminals are mobile across neighborhoods for more serious economic crimes, but that petty crimes and vandalism do not involve any mobility.

JEL Classification : K42, R23, J64

Key words: crime victimization, neighborhood effects, unemployment

---

\*Paris School of Economics

Contact: [camille.hemet@psemail.eu](mailto:camille.hemet@psemail.eu) - Webpage: <http://sites.google.com/site/camillehemet>

I thank Yann Algan, Pierre-Philippe Combes, Bruno Decreuse, Roberto Galbiati, Emeric Henry, Thierry Mayer, Daniel Montolio, Christian Schluter, Olmo Silva, Matthieu Solignac, Grigorios Spanos, Maxime To, Tanguy van Ypersele and Yves Zenou for very useful comments and suggestions. I am grateful to Marc Sangnier for his help with the neighborhoods adjacency matrix. I also thank participants at various seminars (Sciences Po Econ Lunch, Copenhagen Business School, Spatial Economics Research Center -LSE, IZA, U. of Groningen, BETA - U. of Nancy, ULB - Charleroi, U. of Bern, French Ministry of Interior) and conferences (AMSE PhD workshop, III Workshop on Urban Economics (IEB), ADRES doctoral conference in Paris Dauphine French Economic Association (AFSE) annual conference in Lyon and annual meeting of the Urban Economic Association in Washington, LAGV days) for relevant comments. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'avenir" program (reference: ANR-10-EQPX-17 - Centre d'accès sécurisé aux données - CASD), and by three other public grants overseen by the ANR (references: Labex OSE ANR-10-LABX-93-01, ANR-16-FASI-0001-01 and ANR-18-CE22-0013-01).

# 1 Introduction

Some factors such as a high population density or a large unemployment rate, are known for rising crime rates. Living in a deprived US county, Italian province or French department hence puts one at a higher risk of being victim of a criminal event than living in a prosperous region. Yet, such a statement may hide important spatial disparities: a region characterized by well-defined social and economic attributes generally encompasses very heterogeneous areas. As we zoom in and focus on smaller and smaller areas, the relationship between social, economic or demographic characteristics and crime rates established at more aggregate levels may be altered. Consider for instance two adjacent neighborhoods, a prosperous one and a depressed one. Admittedly, unemployment breeds crime, so that the depressed neighborhood will be a nest for criminals.<sup>1</sup> Those criminals may act in their own neighborhood, but they may as well travel to the more attractive adjacent neighborhood. This could lead to a dual crime market, with different types of crimes committed in different neighborhoods (*e.g.* vandalism in poor neighborhoods and burglaries in wealthy areas). This simple example illustrates the idea that studying crime from a more disaggregate perspective can challenge some of the established results, and lead to a better understanding of the mechanisms behind criminal events. In particular, adopting a local approach allows the researcher to ask or revisit the following questions. Does unemployment increases crime? To what extent is the probability to be victim of a criminal event in a given neighborhood affected by unemployment levels in the surrounding areas? Can we observe a duality in crimes, with some crimes explained by intrinsic neighborhoods characteristics and others explained by the characteristics of more distant areas?

This paper answers these questions taking advantage of the French victimization survey that provides detailed information localized at a very low geographic level (a 2,000 inhabitants neighborhood). Note upfront that the survey asks whether individuals have been victims of any criminal event, that is *victimized*, but does not inform about people committing crimes.<sup>2</sup> Hence, I am able to characterize the circumstances of a crime and the victim, but not the criminal. Two important findings emerge from this study. First, among the various neighborhood characteristics considered, unemployment rate appears as the most relevant fac-

---

<sup>1</sup>I use the term "crime" in a very broad sense, encompassing various types of offenses, from misdemeanor to felony. In this respect, "criminals" can designate simple delinquents.

<sup>2</sup>I will henceforth use the term *victimized* to refer to someone having been victim of any criminal event (from property to violent crimes and vandalism). Similarly, a *victimization* will refer to the criminal event making one a victim.

tor having a positive effect on victimization. Second, going beyond the local neighborhood and looking at the surrounding neighborhoods, I find that the effect of unemployment rate in surrounding neighborhoods is stronger than the effect in the reference neighborhood for crimes such as burglaries and thefts of objects from cars, while the reverse is true for smaller crimes.

The present work differs from the previous literature by exploring the determinants of victimization at a very low geographic level. The geographic unit considered, called IRIS, is a 2,000 individuals neighborhood and is the smallest census tract unit for which representative indicators can be constructed in France. Instead, existing results are generally obtained using more aggregate data: [Gould et al. \(2002\)](#) and [Kelly \(2000\)](#) rely on US counties, which add up to 3,140 units for the whole country; [Buonanno et al. \(2009\)](#) and [Bianchi et al. \(2012\)](#) are based on 95 Italian provinces; [Machin and Meghir \(2004\)](#) rely on 43 police force areas for England and Wales; and [Fougère et al. \(2009\)](#) use data from the 96 French *départements*. An exception is [Bell et al. \(2010\)](#) who study the impact of immigration on crime using data from 371 local authorities across England and Wales. Although I am not questioning the validity of the results based on aggregate data, I think that they present an important drawback. As a consequence, these studies ignore the heterogeneity across smaller locations (municipalities, neighborhoods) within the spatial unit of analysis, and thus miss part of the mechanisms through which economic conditions can relate to crime. By contrast, I argue that according to the type of crime and the expected pay-off, criminals might either operate in their own neighborhood or travel to a remote area. It is for instance reasonable to think that thieves are more likely to live in deprived neighborhoods and to steal from wealthier (possibly neighboring) areas, while young delinquents will not have any incentives to bear transportation costs in order to vandalize cars in a distant neighborhood. [Zenou \(2005\)](#) provides some theoretical background for this idea in an urban economics model explaining the link between crime and location by highlighting the role of the housing market. In particular, distance to the city center (where jobs and crime opportunities are located) affects the decision to commit crime instead of working by increasing commuting costs and reducing housing rents. The idea that distance and mobility matter in criminal decision also finds some empirical support in the criminology literature. It documents that the places where perpetrators commit crimes often differ from their area of residence, and that the distance between the two locations varies with the accessibility of the target area, the type of crime and the offender's characteristics (see [Bruinsma \(2007\)](#) for a detailed survey on the Netherlands, and [Bernasco and Luykx](#)

(2003) for an analysis of criminals' target location choice). Working at a very local level enables me to add a spatial dimension to the study of crime, which is a key input to the literature. I am indeed able to compare the effects of the characteristics of adjacent neighborhoods on crime and therefore to capture more precisely the mechanisms behind the relationships obtained with aggregate data.

In addition to allowing for the localization of the surveyed individuals at the neighborhood level, the victimization survey data used in this paper present several valuable features. First, in some cases, it is possible to know where the victimization took place, and hence to control for the characteristics of this location. By contrast, studies relying on police data usually consider the location of the police station where the crime was reported rather than the location of the event itself. Second, these data provide detailed information on individuals, so that relevant individual characteristics pertaining to potential victims' attractiveness can be taken into account, while they are ignored in most of the existing literature. Finally, victimization surveys are known for avoiding the under-reporting issue from which reported police data suffer. Not only are individuals less likely to report personal offenses or small property crimes to the police, but criminal attempts or threats are also not always taken into account by police forces. Relying on victimization survey data is thus particularly insightful regarding petty crimes and assaults. Distinguishing between petty crimes such as vandalism and more important economic crimes shows quite relevant, as these different types of crimes turn out to be driven by different channels.

The nature and the quality of these data enable me to answer the questions asked above. Regarding the local determinants of crime, I show that social, economic and demographic neighborhood characteristics strongly matter to explain victimization. In particular, among the various neighborhood characteristics considered, unemployment rate appears as the most relevant factor, while factors such as the share of immigrants in the neighborhood are not important in explaining victimization. The subsequent results of this paper hence focus on the role of unemployment on victimization. The coefficient for local unemployment rate is positive and its magnitude is particularly strong for small crimes such as motorbike theft or vandalism. Therefore, it seems that crimes committed in more deprived areas relate more to social disorganization theory (e.g. [Shaw and McKay, 1942](#)) than to rational economic crime theory *à la* Becker. Note that these results are obtained after correcting for the biases related to endogenous sorting, as will be explained below.

Finally, following the intuition that perpetrators may move across neighborhoods to commit economic

crimes, I adopt a spatial approach that consists in controlling for both reference neighborhood and adjacent neighborhoods characteristics. The results show that for crimes such as burglaries and thefts of objects from cars, the effect of unemployment rate in distant neighborhoods is stronger than the effect in the reference neighborhood, while the reverse still holds for smaller crimes. Otherwise stated, for a given local unemployment rate, being surrounded by higher unemployment areas increases the probability of being burgled, but does not affect vandalism. Rather, vandalism is boosted by larger local unemployment rates for a given level of unemployment in the surrounding neighborhoods. This tends to support the idea of criminals mobility for some types of crime, e.g economic crimes, in line with Becker's theory, but not for other types of crimes (petty crimes and vandalism), relating instead to the social disorganization theory. This result is, to my opinion, the most important finding of this work. It helps understand the mechanisms behind the finding that, at larger geographic level, unemployment increases crime: unemployment would have a direct local effect on small crimes versus a remote effect on more serious economic crimes. Not only does it mean that the relationship between unemployment and crime is not trivial as we focus on smaller areas, but that this relationship also depends on the type of offense. This result shows the importance of taking criminals' mobility into account, and implies that distance, geography and transport infrastructure might be worth getting more attention in future research on crime.

This paper obviously relates to the large economics of crime literature, initiated by [Becker \(1968\)](#) and [Ehrlich \(1973\)](#). The hypothesis developed in these seminal papers is that the decision to engage into criminal activities is the result of a rational cost-benefit analysis. Most empirical research on the economics of crime aims at testing this hypothesis, which implies that economically weaker individuals (e.g. unemployed workers) have a higher propensity to commit crime because they face lower opportunity costs. Part of the literature hence focuses on economic factors, revealing that lower wages ([Gould et al., 2002](#)), larger unemployment rates ([Fougère et al., 2009](#)) or more inequality ([Kelly, 2000](#)) generate higher crime rates. Alternatively, several studies concentrate on demographic factors such as population density: [Glaeser et al. \(1996\)](#) show that crime is rife in denser and more populated areas due to extended social interactions. A similar idea is developed by [Calvó-Armengol et al. \(2007\)](#) and [Patacchini and Zenou \(2008\)](#) who show the importance of social relationships, in particular of weak ties,<sup>3</sup> in criminal behavior. A growing literature also focuses on the role

---

<sup>3</sup>Weak ties are simple acquaintances, doing contrast to strong ties which are usually close friends and close relatives

played by immigration, and provides evidence that its causal impact on crime is not significant or only very moderate. For instance, [Bianchi et al. \(2012\)](#) demonstrate that the share of immigrants in Italian provinces has only a marginal effect on crime rates through robberies. Other studies, such as [Spenkuch \(2010\)](#) or [Bell et al. \(2010\)](#) show that the effect is driven by the most economically deprived immigrants. On another aspect, [Buonanno et al. \(2009\)](#) insist on the role of social norms and show that they tend to reduce property crimes.

Incidentally, the low geographic focus of this study binds it to the literature on neighborhood effects. A major concern in this literature is that households usually sort across neighborhoods in a non-random fashion. It is then possible that some unobserved household or individual characteristics influence both the propensity to be victim of a criminal event and neighborhood characteristics, therefore biasing the results. Several methods have been used in the literature to overcome this endogenous sorting issue, such as randomized experiments or instrumental variables, that will be detailed more carefully in the paper. The approach adopted in this study follows [Bayer et al. \(2008\)](#) and builds on the very local nature of the data. The idea is that although households are able to select a given area in which they want to live, they are, however, unable to select a precise neighborhood within this given area. Therefore, once the characteristics of the larger selected area are controlled for, the remaining variation of unemployment rates across the smaller neighborhoods can be considered as exogenous.

The rest of the paper is organized as follows. I describe the victimization survey and explain the particular geographic structure of the data in Section 2. In Section 3, I present the strategy used to bypass the issue related to endogenous location selection, and propose several arguments and tests of the identifying assumption. The baseline empirical specification and the corresponding results are presented in Section 4. Section 5 is then devoted to studying the role of unemployment not only in the reference neighborhood, but also from surrounding areas. Section 6 concludes.

## 2 Data overview

The French victimization survey (*Cadre de Vie et Sécurité* - Living Environment and Security, INSEE, CVS henceforth) is a repeated cross section, representative of mainland France households. It has been conducted annually since 2007 and each wave contains approximately 16,000 observations (one per household). The

latest wave I use is the 2011. For each type of victimization considered in the CVS survey, the respondent is asked whether it occurred at least once over the two years preceding the survey. Various types of victimization affecting households in general are considered. These are mostly property thefts (or attempts) and acts of vandalism: burglary, attempt of burglary or theft without breaking in the main home (*burglary*), car theft or attempt, (*car theft*), motorbike (or scooter) theft or attempt (*motorbike theft*), *bike theft*, act of vandalism on the main home (*home vandalism*), act of vandalism on the car (*car vandalism*), and theft of objects from the car (*car objects theft*). I will henceforth refer to these types of victimization as household victimization. In addition, one randomly selected individual in each household is asked about his/her personal experience of victimization over the past two years.<sup>4</sup> In this paper, I will consider three types of individual victimization: *robbery*, *theft* and *assault*.<sup>5</sup> The shares of households and individuals victims of a given type of victimization at least once over the previous two years are displayed in the first column of Table 1. These figures are obtained pooling the 2007 to 2011 waves of the victimization survey. The other columns report the figures for various types of urban units, according to their population size and their degree of urbanization. Expectedly, the probability of victimization is higher in larger urban units (more than 50,000 inhabitants) and in the Paris urban unit than in less populated and rural areas. It is also clear from this table that occurrences of victimization are very rare events, which does not ease their study. The survey reveals that very few households or individuals report repeated occurrence of a given type of victimization, so that considering the occurrence of an event or its number does not make a large difference (this is not in the table).

When a victimization is reported, the respondent gives details about the circumstances, declaration to the police, consequences (physical injuries, protection behavior), and offender (e.g. when s/he was seen or arrested). The data also contain detailed information on households such as income, home ownership status or number of children, as well as individual characteristics such as age, gender, socio-economic category, education, income and national origins. Descriptive statistics of household and individual characteristics are reported in Table 2. In addition, the survey describes the neighborhood: the pollster characterizes the type of housings in the neighborhood and indicates whether s/he observes evidence of vandalism (burnt cars for instance). The respondent also reports whether s/he was aware of any crime or alcohol or drug related incident in the neighborhood and characterizes the general quality of the neighborhood (street light, green

---

<sup>4</sup>The member of the household selected to answer to the individual part of the survey is the person above 14 years old whose birthday is the closest to the 1st of January.

<sup>5</sup>The survey also informs about threats or insults, but I decide to let these types of victimization aside.

spaces, buildings aspect, bunch of people hanging around). Finally, the respondent rates her/his feeling of insecurity.

All this information is available in the public version of the survey. I also have access to more sensitive information, through a Secure Remote Center of Access to the Data (*Centre d'Accès Sécurisé Distant*, CASD). In particular, I am able to localize the precise neighborhood where the surveyed households live. This local area, called IRIS (*Ilots Regroupés pour l'Information Statistique*) is the smallest census tract unit for which representative indicators can be constructed in France. All French municipalities with more than 10,000 inhabitants and most of the municipalities with more than 5,000 inhabitants are divided into several IRISes. Each IRIS is defined so as to be an homogeneous area in terms of living environment, and its borders follow the main topographical and landscape frontiers (e.g. roads, railways and rivers). The target size of an IRIS is 2,000 inhabitants, so that IRISes actually include between 1,800 and 5,000 inhabitants.<sup>6</sup> To give an idea of the level of aggregation, there are about 50,000 IRISes in France (for around 36,000 municipalities). By comparison, there are 96 *départements* in France, the geographical unit used by Fougère et al. (2009).<sup>7</sup>

For the sake of illustration, Figure 1 shows a map of two cities. Paris is represented on the top left panel, and the top right panel focuses on the 18<sup>th</sup> district (*arrondissement* in French), which is colored in darker grey on the map. To give an idea, Paris is divided into 992 IRISes for about 2.2 millions inhabitants, and the 18<sup>th</sup> district comprises 75 IRISes where around 200,000 people lived in 2011. The same year, around 114,000 inhabitants lived in the city of Rouen, in Normandy, in one of 46 IRISes. The surface area of the 18<sup>th</sup> district of Paris is of 6 km<sup>2</sup> while that of Rouen is of 21 km<sup>2</sup>, so that Paris 18<sup>th</sup> district is more than six times denser than Rouen. As a consequence, IRISes in Paris are quite small compared to other cities. In the remainder of the paper, I will interchangeably refer to IRIS or neighborhood.

---

<sup>6</sup>The IRISes are thus comparable in terms of population size, but not necessarily in terms of geographical space. Typically, a small village in the countryside is not divided into IRISes and is actually an IRIS of its own, while cities with more than 5,000 inhabitants are divided into several IRISes. The denser the city considered, the smaller the size of its IRISes.

<sup>7</sup>Bianchi et al. (2012) rely on Italian Provinces, that adds up to a total of 95 units, and most studies on the US are done at the county level, that adds up to 3,140 units.

Figure 1: Neighborhoods divisions (IRIS) in two cities: Paris (left top panel – with a focus on the 18<sup>th</sup> arrondissement, right top panel) and Rouen (bottom panel)



Because each wave of the CVS survey comprises about 16,000 observations, there are very few observations in each IRIS (2.3 observations per IRIS per year on average). Working at such a small scale thus prevents me from computing representative victimization rates at the IRIS level. Instead, I use variables indicating whether each individual or household has ever experienced victimization over the past two years. On the bright side, working at the IRIS level presents a major advantage: it enables me to supplement the victimization data with social, economic and demographic characteristics representative of the IRIS. Indeed, the INSEE designed the IRIS to be the primary statistical unit of the census. Most of the French statistical data sources are therefore based on this geographical unit, so that it is easy to match information at the IRIS level. Using the French population censuses from 2006 to 2009, I can enrich the CVS survey data with socio-economic and demographic characteristics of the IRIS, at the time of the survey. Since 2004, the population census has been conducted annually, in a continuous way, and each wave contains information collected over five consecutive years.<sup>8</sup> For instance, the 2006 census was conducted over the 2004 to 2008 period. Individuals living in municipalities of less than 10,000 inhabitants are all surveyed once over the period. For

---

<sup>8</sup>Prior to 2004, the population census was conducted every decade on average, the latest one dating back from 1999.

municipalities of more than 10,000 inhabitants, 8 % of the population is surveyed each year, so that 40 % of the population is included in the final census data. Because the CVS survey data of a given year concern events that happened over the previous two years, I match them with the census data of the previous year, to be as close as possible in terms of dates: the 2007 wave of the CVS survey is hence matched with the 2006 census data and so on. To be more precise, I enrich the CVS survey data with the following characteristics, representative at the IRIS level: unemployment rate, share of single-parent households, share of immigrants, share of public housing units, share of households arrived less than two years ago and share of 14-18 years old. Furthermore, I can retrieve the IRIS median household income (per consumption unit) from tax surveys. The median household income of a given IRIS is averaged over two consecutive years (weighted by the number of consumption units) and then matched to the following wave of the CVS survey. For instance, the observations from the 2007 wave of the CVS survey are matched with the average median income of 2005 and 2006. Table 3 describes the most relevant contextual variables accounting for households' living environment.

When a victimization is reported in the survey, information is gathered about its circumstances. In particular, the respondent indicates whether it took place in his/her own neighborhood or in some non-specified other place. As I am interested in the local determinants of victimization, I restrict the victimization occurrence to the events that happened in one's own neighborhood, which I am actually able to identify and to characterize. I can then control for the socio-economic environment of the IRIS where the event occurred. Hence, for all types of victimization considered, the dependent variable takes on value 1 if the household or the individual was offended in his own neighborhood and 0 otherwise, i.e. if the offence happened outside the neighborhood or if no offence happened at all. Table 4 documents the extent to which victimization happens in the neighborhood. It shows that most of the household victimization happens in the neighborhood, while the reverse is true for individual victimization. Limiting the study to victimization that happened in the victim's neighborhood can hence be an issue for individual victimization, but it is the only way to control for contextual characteristics.

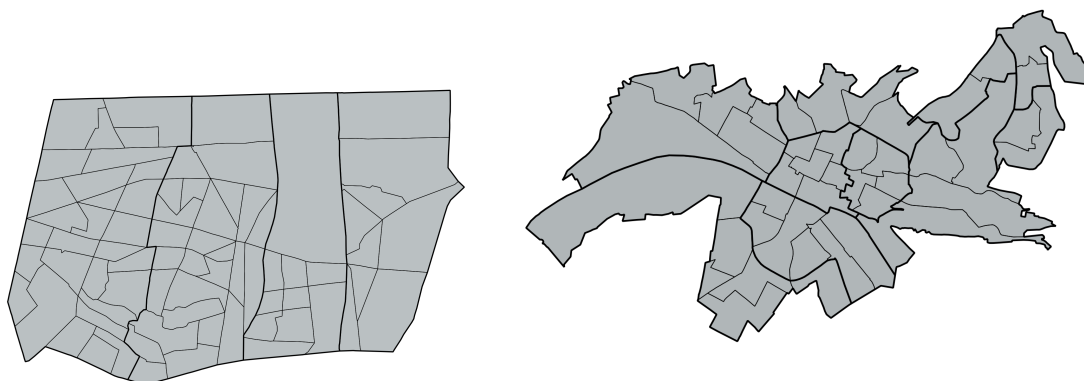
### 3 Addressing the issue of endogenous location selection

A major concern when working with local issues, pervasive in the literature on neighborhood effects, is that households usually sort across neighborhoods in a non-random fashion. It is then possible that some unobserved household or individual characteristics influence both the propensity to be victim of a criminal event and neighborhood characteristics, therefore biasing the results. Several methods have been used to overcome this endogenous sorting issue, that I briefly summarize here. A first approach consists in using a measure of the variable of interest aggregated to a higher geographic level as an instrument for this variable. For instance, [Evans et al. \(1992\)](#) instrument neighborhood poverty with metropolitan area poverty. An alternative method is to rely on randomized experiments designed such that the choice of neighborhood is actually exogenous. One of the most famous examples is the Moving To Opportunity program in the US, through which randomly selected households are given housing vouchers, enabling them to relocate in richer neighborhoods. In particular, and related to the topic of the present study, [Ludwig et al. \(2001\)](#) and [Kling et al. \(2005\)](#) use this experiment to examine the role of neighborhood characteristics on juvenile crime. These are, to the best of my knowledge, the only two existing studies researching the impact of neighborhood effects on crime. However, the very particular setting in which these results are derived brings some concern regarding their validity in a more general context. [Bayer et al. \(2008\)](#) review more extensively these alternative methods and discuss their limitations.

The approach adopted in this paper builds on the very local nature of the data. It follows [Bayer et al. \(2008\)](#) who study the role of neighbors on work location, and more recent work by [Hémet and Malgouyres \(2018\)](#) and [Solignac and Tô \(2018\)](#) who investigate neighborhood effects using French data. The idea is that although households are able to select a given area in which they want to live, they are, however, unable to pinpoint a precise neighborhood within this given area. This assumption means that even if households are able to choose a given residential area, there will not be any correlation in unobserved factors affecting risk of victimization among individuals living in the same neighborhood *within the larger selected area*. As a consequence, once we control for the characteristics of the larger area selected by the individual, the remaining spatial variation of characteristics across neighborhoods within the larger area is supposed to be exogenous. Empirically, this is done through the inclusion of fixed effects of these larger areas. In particular, in this paper, I assume that the area into which households select corresponds to what is literally called a *Large*

*Neighborhood* in the French statistical jargon (*Grand Quartier*). Large enough municipalities are divided into several *Large Neighborhoods*, which themselves encompass several contiguous IRISes. As for IRISes, Large Neighborhoods are following obvious physical boundaries (e.g. streets, rivers...) and are designed by the French Statistical Institute so as to be as homogeneous as possible in terms of living environment. Figure 2 depicts the four Large Neighborhoods of Paris 18<sup>th</sup> district and the ten Large Neighborhoods of Rouen, which are delineated by thicker lines. Although there is no formal evidence that large neighborhood is the geographical unit targeted by households when looking for a housing, this area makes a reasonable reference neighborhood compared to the IRIS.<sup>9</sup> Importantly for my argument, these large neighborhoods are generally labeled by a "common-knowledge" name referring to a neighborhood to which people can actually refer (e.g. "city-hall" or "station" neighborhood). In other words, households can easily identify these Large Neighborhoods, by contrast with IRISes as explained below. In the case of Paris, Large Neighborhoods correspond to "administrative neighborhoods" upon which a number of public services are administered. In particular, each of these administrative units possess a police station. Relying on the assumption that households decide upon the Large Neighborhood where they want to live, but do not pick a specific IRIS *within* this Large Neighborhood, I can subsequently estimate local neighborhood effects once including Large Neighborhoods fixed effect: the identification comes from the remaining variation of characteristics across IRISes within a given Large Neighborhood.

Figure 2: Large Neighborhoods in Paris 18<sup>th</sup> (left panel) and Rouen (right panel)



<sup>9</sup>An alternative and maybe more obvious selection zone would be schools' catchment areas. Although it is not easy to get access to the corresponding maps, except for some specific regions, I intend to focus on these areas as an alternative selection zone.

It is now in order to present a few arguments supporting this assumption. First, because the housing market is very tight, it is reasonable to think that a household targeting a given area very unlikely has a choice over the precise neighborhood where it will end up in this area. This would indeed require that at least one housing unit satisfying the other decision criteria of the household (e.g. size) be vacant in each neighborhood within the larger area at the time when the household is looking for a new place. A second consideration is that it may be difficult for prospecting households to identify neighborhood-by-neighborhood variation in neighbors and contextual characteristics, prior to moving into the neighborhood. To put it differently, although the household may have a realistic *ex-ante* view on the characteristics of the targeted area, it is less likely to be actually able to identify differences in these characteristics across the various neighborhoods of the area. This makes even more sense in the context of victimization, for which *ex-ante* information is particularly difficult to gather. Finally, an interesting feature of the French neighborhoods studied here (the IRISes) is that they do not follow any kind of administrative frontier. For instance, they are distinct from police districts, and from school zones determining to which school children must go. Rather, the neighborhoods considered here were designed to encompass 2,000 inhabitants on average, and to be homogeneous in terms of living environment, with borders following the main topographical and landscape frontiers (e.g. roads, railways and rivers). People ignore where these borders are, and more generally do not even know what an IRIS is, as it is only used for statistical purpose. For those reasons, it is practically impossible that households purposely decide to live in a given IRIS rather than in a contiguous one. All these arguments support the validity of the assumption that there should be no correlation in unobserved factors affecting victimization among neighbors living in the same neighborhood (IRIS) *within the larger selected area*.

In addition to these general arguments, I now turn to a more formal statistical test of this assumption: I want to verify whether there is sorting across local neighborhoods based on individual observable characteristics, and whether this sorting remains after controlling for selection into Large Neighborhoods (*i.e.* I want to test for the presence of sorting across IRISes within Large Neighborhoods). The test I propose here is inspired by [Hémet and Malgouyres \(2018\)](#) who rely on a similar strategy but on different spatial units. I first check whether there is sorting across neighborhoods. To this aim, I randomly pick one individual in each IRIS and regress a given observable characteristic (*e.g.* origin or education) on the IRIS-neighbors average of this characteristic, excluding the individual and members of her household. This procedure is repeated 100

times. The average R-squared from these regressions for different observable characteristics are reported in column 1 of Table 5. As expected, they are rather large, suggesting a substantial degree of spatial sorting.

I now want to test the assumption that there is little sorting remaining after including fixed effects for larger selection areas. I first regress any given individual characteristic on fixed effects for larger geographic areas into which households are likely to self-select, from the largest to the smallest (*départements*<sup>10</sup>, municipalities and IRISes), and I store the residuals from these regressions. I then store the residuals obtained after running the same regressions using instead the corresponding neighborhood average as the dependent variable. Both residuals simply measure deviation from the area average. If households do not actually select in which particular IRIS they settle *within* a given larger area, there should be no association between the two residuals. To test this, we regress the residual from the individual regression on the residual from the neighbors-average regression, and we eventually store the the corresponding R-squared. I repeat this procedure 100 times and compute the average R-squared, which are eventually reported in columns 2 to 4 of Table 5. Looking at the R-squared reported in column 4, obtained when fixed effects for Large Neighborhoods are included, we can note that they are much smaller than those reported in the previous columns, and in particular than those obtained in the unconditional regressions in column 1. This supports the assumption that the amount of *within-Large Neighborhood* sorting based on the various observable characteristics is negligible.

Finally, Solignac and Tô (2018) also work at the IRIS level to study how local employment influences labor market outcomes of young graduates, and rely on the same identifying assumption. In their paper, they propose an alternative statistical test of the validity of this assumption that I summarize here. This test consists in showing that the spatial autocorrelation in observed characteristics between local neighborhoods is limited once these local neighborhoods are nested within Large Neighborhoods. To this aim, the authors rely on Moran indices, which are designed to capture spatial autocorrelation in a given variable between all units (here, IRISes) belonging to the same geographical area. This statistic is large and positive if local neighborhoods with similar characteristics tend to be clustered spatially, and it gets close to zero if local neighborhoods are randomly spread out. Using data from the 1999 census, they compute Moran indices for a set of IRIS characteristics, taking alternative larger spatial units as the reference area. For all characteristics considered, the distribution of standardized Moran statistics is much less dispersed when the reference unit

---

<sup>10</sup>*Départements* are large administrative units: there are 96 such units in mainland France.

(i.e. the unit within which spatial autocorrelation is measured) is the Large Neighborhood compared to larger areas such as the municipality. This means that households tend to sort themselves between local neighborhoods within municipalities for instance, but not within Large Neighborhoods where spatial autocorrelation remains quite limited, providing additional support to our common identifying assumption.

## 4 Empirical specification and baseline results

Let  $i$ ,  $j$  and  $k$  indicate respectively individual, household and IRIS. For each outcome considered, I estimate the following equation.

$$VICT_{(i,j),k} = \alpha + \beta U_{k(i,j)} + X_{i,i(j)}\gamma + Y_{j(i),j}\delta + Z_{k(i,j)}\kappa + \mu F_{year} + \nu F_{LN} + \varepsilon_{(i,j),k} \quad (1)$$

where  $VICT_{(i,j),k}$  is a dummy variable indicating the occurrence of a given type of victimization at least once over the preceding two years. More precisely, in the case of a household victimization the dependant variable is  $VICT_{j,k}$ , while it is  $VICT_{i,k}$  when it turns to individual victimization, with  $k$  standing for the IRIS where the surveyed household lives.  $U_k$  stands for local neighborhood unemployment rate, so that  $\beta$  is the main parameter of interest.  $X$  is a vector of individual observables characterizing the surveyed individual or the head of households according to the type of victimization considered (individual or household respectively),  $Y$  is a vector of household characteristics and  $Z_k$  a vector of social, economic and demographic characteristics of the IRIS, along with other contextual variables that are detailed below. I also include year fixed effects ( $F_{year}$ ) to account for specific time trends related to a change in economic conditions or policy orientation for instance. Finally, *Large Neighborhoods* fixed effects ( $F_{LN}$ ) are included for the sake of identification, as explained in the previous section. All results presented below derive from the estimation of a linear probability model, using OLS estimates, with robust standard errors clustered at the IRIS level.

Two broad sets of variables are used in the regressions: one to control for the living environment in a general sense ( $Z_{IRIS}$ ) and another to control for individual and household characteristics ( $X$  and  $Y$ ). Regarding contextual variables, I control for social, economic and demographic neighborhood (IRIS) characteristics: median annual household income (in log), share of immigrants, share of households living in the public housing sector, share of 14- to 18-year-old individuals, share of single-parent families (*monoparental*) and

share of households that have been living in the IRIS for less than two years (*recent movers*). I also include a variable from the CVS survey describing the type of buildings in the neighborhood (dispersed houses out of the city, houses in a lot or in the city, apartment blocks in the city or in the suburbs). Regarding household characteristics, the following controls are used: household monthly income (in three categories), ownership status (owner, tenant in the private housing sector or tenant in the public housing sector) and number of children in the household. As far as individual (respectively household head) characteristics are concerned, age, gender, nationality, occupation status and socio-economic category of the surveyed individual (respectively household head) are included in regressions of individual (respectively household) victimization.

Table 6 summarizes the results of the regressions including large neighborhood fixed effects. The coefficients for the variables other than IRIS characteristics are not presented for the sake of space-saving, but are available upon request, and are commented below.

The estimates reveal that living in a neighborhood with a larger share of immigrants would reduce the probability of being burgled and robbed. This is particularly striking given that people's feeling of insecurity is strongly positively correlated with the share of immigrants in the neighborhood. Living in a poorer area would also lower the likelihood of robberies compared to richer neighborhoods, which goes in the expected direction given the incentives. The share of single-parent households appears to drive car thefts and car vandalism, as well as thefts of objects from cars, but to a lesser extent. The share of recent movers is also an important characteristic explaining vandalism. A possible interpretation is that the larger the share of recent movers, the less neighbors know each other, and hence the less likely they are to organize some sort of collective neighborhood watching. Another interpretation pertaining to the social disorganization theory is that weaker social ties undermine the ability of a community to exercise informal control over its members. This alternative explanation is also relevant to the previous results on single-parents families.

Local unemployment rate emerges as one of the most important determinants of crime victimization, both in terms of the number of victimization types in which it is involved and in terms of magnitude of the coefficient. The effect of unemployment rate is particularly large for car vandalism. Because this is a non-lucrative and violent offense, this result also seems in line with social disorganization theory, resulting from social rather than financial deprivation. The strong relationship of unemployment with burglaries is less intuitive, as it suggests that burglars live and burgle in the same neighborhood. One interpretation could

be that the burglaries happening in high unemployment neighborhoods are committed by amateurs, out of despair and as a last resort, who seek to steal goods for their personal use (e.g. TV sets, video game consoles, food) rather than for reselling them (e.g. jewelry, pieces of art). An alternative explanation can be borrowed from sociology: the routine activity theory (Cohen and Felson, 1979): one's neighbors are easy to observe on a daily basis, up to knowing when they are absent from home and when they will return (the routine), facilitating burglaries.

The type of neighborhood is also particularly relevant for most types of victimization. As expected, households living in residential areas made of groups of houses are more likely to be offended than those living in isolated houses in the countryside. Households living in apartments buildings are less likely to be burgled, but more likely to have their car vandalized, especially if they live in the suburbs.<sup>11</sup> Now taking a quick look at household and individual characteristics, we can see that wealthier households suffer more of home vandalism, but less of car vandalism, while their members are less likely to be assaulted. The result for cars is probably explained by the fact that wealthier household park their cars in a closed or secured space. Households with an unemployed head are more likely to be victims of burglary, home vandalism, and car theft. Similarly, unemployed individuals are more likely to suffer from violent crimes (robberies and assaults). Older individuals tend to be less victimized. Gender does not affect the probability of theft or robbery, but males are more victims of assaults than females.

Although I do not present the naive estimates in the paper (those estimated without including Large Neighborhood fixed effects, controlling only for larger regions), a few remarks are in order. To a very large extent, the coefficients on IRIS characteristics are smaller once *Large Neighborhoods* fixed effects are included. In particular, for the share of immigrants, the coefficients for *theft of objects from cars* and *theft* are significantly positive in the naive regression and are now driven down to zero (in part due to an increase in the standard errors), while some of the coefficients that were not significantly different from zero are now significantly negative (*burglary* and *robbery*). This suggests that controlling for large neighborhood fixed effects actually corrects a bias induced by the fact that immigrants tend to settle in more criminogenic areas, because of lower rents for instance. The estimates for unemployment are rather stable for burglary, car vandalism and assault. Yet, they are driven down to zero for car theft, motorbike theft and bike theft once we control for *Large Neighborhoods* characteristics. The exception is for robbery, where it becomes slightly

---

<sup>11</sup>They are also less likely to have their home vandalized, but this is mechanically due to their home type.

positive.

## 5 Going beyond one's own neighborhood

An important dimension to take into account in the study of crime is criminals' mobility. To put it simply, if criminals are not mobile, the larger the number of criminals living in a neighborhood, the more likely the other inhabitants of this neighborhood experience victimization. On the other hand, if criminals are mobile, then even individuals living in a criminal-free neighborhood may face a risk of victimization if they are located at some reasonable distance of a neighborhood populated with criminals. Consider for instance a high unemployment neighborhood, more likely to breed criminals according to Becker's theory. The potential offenders could commit crime in the neighborhood where they reside, if, for instance, they cannot afford the cost of commuting to a more distant neighborhood, or if they benefit from observing their neighbors' habits and routine activities (Cohen and Felson, 1979). Alternatively, they could decide to commit a crime in a more distant neighborhood if they fear to be more easily identified in their own neighborhood, or if their neighborhood is too deprived to be attractive. Whether an offender decides to act in his own neighborhood or in a remote one thus reflects a weighting of the expected gains, the direct costs and the opportunity costs of committing crime in another neighborhood. Therefore, even if we find that unemployment increases victimization on average, the effect might actually depend on where one lives relative to high unemployment neighborhoods.

Although the question of criminals' mobility seems highly relevant when studying determinants of victimization, it is not addressed in the economics of crime literature, mainly due to the fact that most studies rely on aggregate data. By contrast, because I work with data localized at a low geographic level, I am able to connect individuals not only to the characteristics of the neighborhood where they live, but also to the characteristics of the neighborhoods that are further away. This new spatial approach enables me to indirectly account for criminals mobility.<sup>12</sup> To this aim, I consider the IRIS where the surveyed individual lives as the reference neighborhood, and I construct two successive circles of adjacent IRISes to represent more distant neighborhoods. More precisely, all the IRISes contiguous to a given IRIS constitute the first ring of adjacent neighborhoods (denoted as *IRIS 1*), while all the IRISes contiguous to those in the first ring, excluding the

---

<sup>12</sup>It is not direct as I have no information about the offenders, so I am not actually able to locate them.

reference IRIS and the first ring of IRISes themselves, constitute the second ring of adjacent neighborhoods (denoted as *IRIS 2*). Figure 3 illustrates the three geographic levels on which I rely for the 18<sup>th</sup> district of Paris (left) and the city of Rouen (right). On each map, the reference IRIS (*i.e.* where the individual lives) is colored in black, while the first and second rings of surrounding IRISes are colored in dark grey and in light grey respectively.

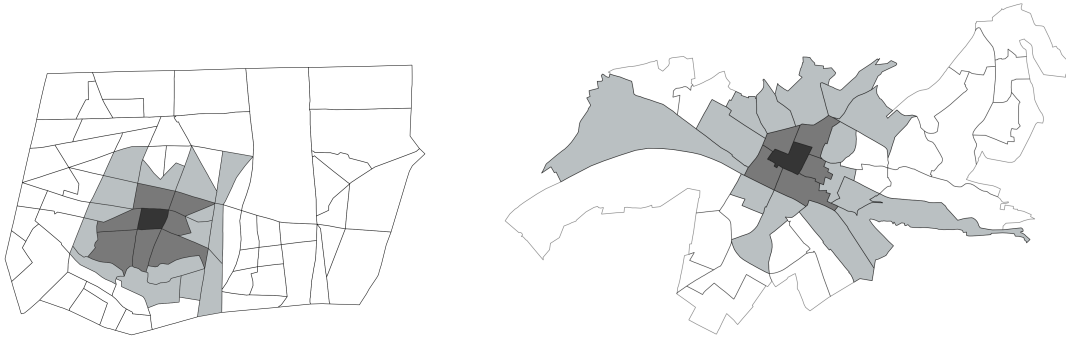
Using this setting, it is possible to relate any individual or household in the survey to the characteristics of the neighborhood where it lives, as well as to those of the first and second rings of adjacent neighborhoods. Thus, I can explore whether a given factor matters more within the neighborhood or from a remote one. In the following empirical analysis, I will focus on one particular factor: the unemployment rate. First, as noted at the end of the previous section, this factor shows the most relevant in explaining victimization. Second, I make this choice in order to avoid a likely collinearity issue with other IRIS characteristics. The unemployment rate of a given neighborhood is indeed highly correlated with most of the other IRIS characteristics (share of immigrants, median income, share of single parents families and share of public housing), as can be seen from Table 7. Concretely, I compute the average unemployment rate over all the first and second rings of IRISes respectively, weighted by the size of the active population in each IRIS. To summarize, and using the same notation as in section 4, I estimate the following equation:

$$VICT_{(i,j),k} = \alpha + \beta_0 U_{k(i,j)} + \beta_1 U_{k+1(i,j)} + \beta_2 U_{k+2(i,j)} + Y_{j(i),j} \delta + Z_{k(i,j)} \kappa + \mu F_{year} + \nu F_{LN} + \varepsilon_{(i,j),k} \quad (2)$$

Where  $U_k$  is the unemployment rate in the reference IRIS,  $U_{k+1}$  the average unemployment rate in the first ring of adjacent IRISes, and  $U_{k+2}$  the average unemployment rate in the second ring of adjacent IRISes.

This geographical approach has some drawbacks. First, as I do not have any information for the road or transportation networks, I am not effectively capturing transportation time or cost, which are key determinants of mobility. This could be addressed using the information about road networks provided by the French Institute of Geography (IGN), and performing a Geography Information System analysis. However, because I do not have access to these data nor to the technology necessary to deal with it, I keep this step for future research. Second, this approach with adjacent IRISes does not enable me to directly capture distance, as IRISes are heterogeneous in terms of size. As mentioned in Section 2, only municipalities with more

Figure 3: Examples of two reference neighborhoods (black) with their first (dark grey) and second (light grey) rings of surrounding neighborhoods, in Paris 18<sup>th</sup> and Rouen.



than 5,000 inhabitants are divided into IRISes, and the target size of an IRIS is 2,000 inhabitants so that denser cities tend to have smaller IRISes. To deal with this issue, I restrict the sample to municipalities that are actually divided into IRISes, hence reducing the variation in the size of the IRISes. Such a restriction typically excludes rural villages, which are quite large IRISes (in terms of surface).

The estimated effects of the unemployment rate in the three successive neighborhood rings (IRIS, IRIS 1 and IRIS 2) on the various types of victimization are reported in Table 8. The first set of results displayed are the estimates obtained when no other control is included, while the second set of results is obtained including the full set of controls. Note that for each specification, I control for large neighborhood fixed effects to avoid the endogeneity issue, as discussed in the previous section. Let us first look at the burglaries in column (1). In the no other control specification, only the unemployment rate in the first ring of adjacent IRISes is positive and significant at the 10% level. In the full specification, both the IRIS and the IRIS 1 unemployment rates are significant at the 5% level, with a larger coefficient for the latter. The results are similar for thefts of objects from cars. The unemployment rate in the first contiguous neighborhoods is the only significant one that in the full specification. This suggests that economic types of crimes such as burglary and theft of objects from cars are better explained by the unemployment rate from more distant places than from the immediate neighborhood. This is in line with the idea that when it turns to economic crimes, offenders are more likely to travel to some remote area. Several considerations can help rationalize this: stealing from one's direct neighbors is not financially attractive when one lives in a more economically deprived neighborhood; the expected financial gain from an economic crime allows the offender to afford

the cost of traveling to a more distant neighborhood; and the criminal limits the chance to be identified by witnesses when committing an offense in a different place than the one where he lives. It is then a bit puzzling that the unemployment rate in the reference IRIS still matters for burglaries. A possible explanation is that there are two types of burglars: those who travel to a remote place to steal expensive goods they can resell such as jewelry and works of art, and those who steal very basic goods such as food or TV sets from their own neighbors for their personal consumption. An alternative explanation could be that the habits and general behavior of one's direct neighbor are more easily observed, so that it simplifies the planning of the crime (Cohen and Felson, 1979). It is also surprising not to find any significant effect on car and motorbike thefts. A possible explanation could be that stealing this type of goods requires an even longer distance, so that it is more easy to stock or use the car.

On the other hand, Table 8 also shows that the unemployment rate in the immediate neighborhood is particularly relevant in explaining non-economic and violent crimes such as acts of vandalism, whether on the home or on the car, and assaults. In this case, the social disorganization theory is more appropriate to understand the mechanisms than the Beckerian approach.

The concern linked to the IRIS size may persist, with the existence of very small IRISes in densely populated cities such as Paris. In this case, the distance between two IRISes may not be relevant, with a null transportation cost across the three contiguous rings of IRISes. In what follows, I therefore exclude the observations of the three largest cities (Paris, Lyon and Marseille), hence getting rid of the smallest IRISes. The regressions presented above are then replicated on this sub-sample. The estimates for unemployment rates in the successive rings of IRISes in the full specification are reported in Table 9. The previous results are robust to this sample restriction. The results are stable for burglaries, with a positive effect of direct (IRIS) unemployment rate and a larger positive effect of more distant (IRIS 1) unemployment. Note that the gap between the two coefficients is even slightly larger than in the previous table. The estimates for both acts of vandalism, theft of objects from car and assaults are also similar to those presented above. There are however two differences compared to the regressions including large cities: the coefficient for IRIS unemployment rate is now significantly positive (at the 10% level) in the motorbike theft and in the robbery equation. To the extent that robberies are violent crimes by opposition to thefts, this new result tends to comfort the idea that the direct exposure to unemployment affects violent rather than economic crimes. The result for motorbike

theft could be at odds with this intuition, unless most of the thefts observed apply to motorcycles rather than to more powerful motorbikes.

Because the unemployment rates of the adjacent local neighborhoods are averaged over the first and second rings of adjacent IRISes, this measure of unemployment may fail to capture important spatial disparities in terms of economic conditions. For instance, a given average unemployment rate may be obtained with a set of IRISes characterized by the same unemployment rate, or with a set of IRISes in which some have a very low unemployment rate and some a very large one. Thus, a given average unemployment rate may hide very different situations. Even if my level of analysis is quite small, it is possible that a given IRIS has very heterogeneous neighbors, so that looking at mean unemployment rates might not be the most relevant metrics. I therefore estimate the same equation as above, but considering the largest unemployment rate among the various IRISes in the first and second rings, rather than the average.

The corresponding estimates are displayed in Table 10, where estimates in Panel A are obtained using the main sample, while those in Panel B are obtained excluding Paris, Lyon and Marseille. The main results are robust to this change in specification.

## 6 Conclusion

This paper is, to the best of my knowledge, the first study on victimization at the neighborhood level. This local approach brings new insights to the economics of crime literature as it enables me to characterize precisely the context (both location and victim) in which criminal events occur. By contrast with previous papers based on aggregate police data, I am therefore able to distinguish between factors related to the opportunity cost of committing crime (e.g. unemployment, wages) and factors pertaining to the attractiveness of the victims (e.g. wealth). I find that household and individual characteristics are minor determinants of household and individual victimization respectively, while the economic situation of the neighborhood actually matters. In other words, victims (individuals or households) are not directly targeted (except in the case of assaults): rather, it is the neighborhood where the mischief occurs that is coveted. In particular, local unemployment rate is found to be strongly related to household victimization. In order to address the endogenous neighborhood selection issue, I included "large neighborhood" fixed effects in order to control

for the characteristics of the larger area that the households are likely to have actually selected. Most of the estimates of neighborhood characteristics are attenuated once selection is corrected for. Yet, the local unemployment rate remains a strong predictor of several types of victimization.

This paper also sheds new light on the mechanisms behind this relationship, through the adoption of an original spatial approach. I take advantage of the precise location of the data to control for the characteristics of both the reference neighborhood and the first and second rings of adjacent neighborhoods. That way, I can account for heterogeneity across neighborhoods and hence indirectly for criminals mobility. This is an improvement over the existing literature which ignores this dimension. The results reveal that for burglaries and thefts of objects from cars, unemployment rate in the adjacent neighborhoods have a stronger explanatory power than unemployment in the precise neighborhood where the misdeed occurred. On the contrary, local unemployment rate dominates over more distant unemployment rates in explaining vandalism and assaults in particular. A natural interpretation of these findings is that criminals are mobile for economic crimes but not for violent crimes. In other words, they can afford some transportation cost when they expect a financial reward from their mischief, in line with the Beckerian theory of crime. On the other hand, violent crimes and vandalism escape from this logic and relate more to the social disorganization theory. This new method helps understanding more precisely criminal behavior according to the different types of crimes, and is therefore a key contribution to the literature.

Naturally, the empirical design endorsed in this paper presents some drawbacks and will be subject to future improvements. For instance, considering only two rings of adjacent neighborhoods is somehow arbitrary and is an important limitation as criminals may travel from more remote areas. In particular, car thefts may involve longer distances so as to reduce the risk of apprehension. This could explain why none of the unemployment rate estimates (IRIS, IRIS1 and IRIS2) is significant for this type of crime. One of the next developments of this work will therefore be to take into account all neighborhoods in an exhaustive fashion. The idea would be to express crime as a function of the sum of unemployment rates in all surrounding neighborhoods, weighted by distance or transportation costs. In other words, this would consist in adapting the market-potential function developed in the new economic geography literature (e.g. [Harris, 1954](#); [Hanson, 2005](#)) to the economics of crime literature. Because it reveals the relevance of a spatial approach and stresses its necessity, the present paper is a first step in this direction.

Table 1: Share of households or individuals victimized at least once over the past two years.

		Full Sample (1)	Rural Areas (2)	Less than 50,000 inhab. (3)	More than 50,000 inhab. (4)	Paris Urban Unit (5)
<b>Household victimization</b>						
Burglary	Mean	4.61 %	3.81 %	4.01 %	5.20 %	5.18 %
	StDev	(.210)	(.191)	(.196)	(.222)	(.222)
	N	85141	16211	19175	35755	13895
Car Theft	Mean	3.36 %	1.59 %	2.97 %	4.19 %	4.87 %
	StDev	(.180)	(.125)	(.179)	(.200)	(.215)
	N	69226	14953	16599	28183	9413
Motorbike Theft	Mean	5.34 %	2.25 %	4.42 %	7.28 %	8.91 %
	StDev	(.225)	(.148)	(.205)	(.260)	(.285)
	N	10051	2633	2470	3755	1181
Bike Theft	Mean	3.71 %	1.14 %	2.56 %	5.52 %	6.34 %
	StDev	(.189)	(.106)	(.158)	(.228)	(.244)
	N	46321	10974	11687	17877	5730
Home Vandalism	Mean	4.13 %	1.93 %	3.93 %	5.68 %	3.69 %
	StDev	(.199)	(.138)	(.194)	(.231)	(.188)
	N	85142	16214	19177	35751	13895
Car Vandalism	Mean	10.46 %	5.66 %	8.46 %	13.42 %	14.48 %
	StDev	(.306)	(.231)	(.278)	(.341)	(.352)
	N	69192	14955	16593	28170	9396
Car Object Theft	Mean	6.71 %	3.52 %	5.41 %	8.16 %	10.79 %
	StDev	(.250)	(.184)	(.226)	(.274)	(.310)
	N	69227	14953	16598	28186	9412
<b>Individual victimization</b>						
Robbery	Mean	0.95 %	0.29 %	0.54 %	1.13 %	2.09 %
	StDev	(.097)	(.054)	(.073)	(.106)	(.143)
	N	85154	16213	19177	35759	13900
Theft	Mean	3.38 %	2.46 %	2.88 %	3.74 %	4.62 %
	StDev	(.181)	(.155)	(.167)	(.190)	(.210)
	N	85148	16211	19176	35759	13897
Assault	Mean	2.42 %	1.76 %	2.03 %	3.09 %	2.34 %
	StDev	(.154)	(.132)	(.141)	(.173)	(.151)
	N	85142	16212	19171	35758	13896

Table 2: Sample characteristics: households and individuals characteristics.

<b>Household Characteristics</b>				
	[Min - Max]	Mean	(StDev)	Med
<i>Household monthly income:</i>				
$w \leq 1500$	[0 - 1]	0.313	(0.464)	0
$1500 < w \leq 2500$	[0 - 1]	0.294	(0.456)	0
$w > 2500$	[0 - 1]	0.393	(0.488)	0
<i>Ownership Status:</i>				
Owner	[0 - 1]	0.598	(0.490)	1
Rent in private market	[0 - 1]	0.213	(0.410)	0
Rent in public housing	[0 - 1]	0.144	(0.351)	0
Other	[0 - 1]	0.045	(0.207)	0
<i>Household composition:</i>				
Head with a partner	[0 - 1]	0.569	(0.495)	1
Number of children	[0 - 11]	0.644	(0.993)	0

	<b>Household Head</b>				<b>Individual</b>			
	[Min - Max]	Mean	(StDev)	Med	[Min - Max]	Mean	(StDev)	Med
<i>Age</i>	[15 - 101]	53.25	(17.87)	52	[14-102]	47.19	(19.62)	46
<i>Gender</i>	[0 - 1]	0.622	(0.485)	1	[0 - 1]	0.479	(0.500)	0
<i>Nationality:</i>								
Native French	[0 - 1]	0.908	(0.290)	1	[0 - 1]	0.907	(0.290)	1
Naturalized French	[0 - 1]	0.043	(0.203)	0	[0 - 1]	0.042	(0.201)	0
EU 15	[0 - 1]	0.021	(0.142)	0	[0 - 1]	0.019	(0.136)	0
Other EU (after 2004)	[0 - 1]	0.001	(0.037)	0	[0 - 1]	0.001	(0.037)	0
Maghrebian	[0 - 1]	0.013	(0.115)	0	[0 - 1]	0.014	(0.118)	0
Other African	[0 - 1]	0.005	(0.073)	0	[0 - 1]	0.006	(0.075)	0
Other nationality	[0 - 1]	0.009	(0.092)	0	[0 - 1]	0.010	(0.100)	0
<i>Employment status:</i>								
Employed	[0 - 1]	0.559	(0.496)	1	[0 - 1]	0.488	(0.500)	0
Unemployed	[0 - 1]	0.041	(0.198)	0	[0 - 1]	0.059	(0.235)	0
Inactive	[0 - 1]	0.340	(0.490)	0	[0 - 1]	0.453	(0.498)	0
<i>Socio-economic Category:</i>								
Farmer	[0 - 1]	0.014	(0.117)	0	[0 - 1]	0.012	(0.108)	0
Craftsman, shopkeeper	[0 - 1]	0.046	(0.209)	0	[0 - 1]	0.034	(0.182)	0
Higher occupation	[0 - 1]	0.111	(0.314)	0	[0 - 1]	0.078	(0.269)	0
Intermediate occupation	[0 - 1]	0.150	(0.357)	0	[0 - 1]	0.128	(0.334)	0
Employee	[0 - 1]	0.130	(0.336)	0	[0 - 1]	0.165	(0.371)	0
Factory worker	[0 - 1]	0.155	(0.362)	0	[0 - 1]	0.131	(0.337)	0
Retired	[0 - 1]	0.353	(0.478)	0	[0 - 1]	0.286	(0.452)	0
Other inactive	[0 - 1]	0.041	(0.198)	0	[0 - 1]	0.166	(0.372)	0

Reading: The head of the average household is 53 years and 3 months old. The average surveyed individual is about 47 years and 2 months old. 55.9 % of household have an employed head. 48.8 % of individuals are employed.

Table 3: Sample characteristics: households living environment

<b>Contextual Variables:</b>	[Min - Max]	Mean	(StDev)	Med
<i>IRIS Characteristics:</i>				
Share of immigrants	[0 - 0.794]	0.079	(0.074)	0.055
Median income (log)	[7.69 - 10.98]	9.792	(0.258)	9.784
Unemployment rate	[0 - 0.741]	0.112	(0.056)	0.100
Share single-parent families	[0 - 0.673]	0.137	(0.068)	0.127
Share hh in public housing	[0 - 1]	0.136	(0.187)	0.064
Share of recent movers	[0 - 0.935]	0.129	(0.061)	0.116
Share of 14-18 y.o.	[0 - 0.239]	0.056	(0.017)	0.056
<i>City density (log)</i>	[-1.09 - 10.55]	6.297	(1.955)	6.292
<i>Type of neighborhood:</i>				
Dispersed houses	[0 - 1]	0.176	(0.381)	0
Houses Lot / in cities	[0 - 1]	0.443	(0.497)	0
Apartment block (city)	[0 - 1]	0.231	(0.422)	0
Apartment block (suburbs)	[0 - 1]	0.091	(0.288)	0
Mixed	[0 - 1]	0.059	(0.235)	0
<i>Size of the Urban Unit:</i>				
Rural Areas	[0 - 1]	0.226	(0.418)	0
Less than 50,000	[0 - 1]	0.251	(0.433)	0
More than 50,000	[0 - 1]	0.365	(0.481)	0
Paris Urban Unit	[0 - 1]	0.158	(0.165)	0

Reading: The average household lives in an IRIS where there are 7.9 % of immigrants. 59.8 % of households own their home. The head of household lives with a partner in 56.9 % of households.

Table 4: Probability that the incident occurs in own's neighborhood

	Mean	(StDev)	N
<b>Household Victimization</b>			
Car theft	0.724	(0.447)	2,052
Motorbike theft	0.643	(0.479)	497
Bike theft	0.755	(0.430)	1,620
Vandalism on the car	0.657	(0.475)	6,581
Theft of object from car	0.289	(0.454)	4,080
<b>Individual Victimization</b>			
Robbery	0.396	(0.489)	633
Theft	0.301	(0.459)	2,308
Assault	0.372	(0.484)	1,725

When at least one offence is reported, more details are asked about the latest event. In particular, the respondent indicates whether the incident happened in one's "own village or neighborhood". Reading: 72.4 % of the latest car theft happened in the owner's neighborhood. 37.2 % of the latest assaults occurred in the victim's neighborhood.

Table 5: Sorting test based on observable characteristics

Variable	(1)	(2)	(3)	(4)
	Unconditional	Département FE	Municipality FE	Large N'hood FE
North African	5.5 %	4.5 %	2.9 %	1.6 %
Master, PhD	10.4 %	5.3 %	2.5 %	1.3 %
Graduate	0.56 %	0.43 %	0.26 %	0.17 %
Undergraduate	0.07 %	0.03 %	0.02 %	0.01 %
Lower undergraduate	0.11 %	0.07 %	0.04 %	0.03 %
Technical Baccalaureate	1.54 %	0.55 %	0.15 %	0.07 %
Lower degree	0.67 %	0.25 %	0.11 %	0.06 %
Drop-out	5.6 %	4.5 %	3.03 %	1.8 %
Age	3.65 %	2.82 %	1.89 %	1.48 %

In each IRIS, we randomly draw an individual and compute the average characteristics among her neighbors (excluding members of her own household). For each observed characteristic, we regress the individual's observation on IRIS-neighbors' average observation and look at the within-group R-squared of the regression. We repeat this process 100 times and compute the average R-squared, which are reported in the table. Column 1 displays the average R-squared from the unconditional regressions, while Columns 2, 3, and 4 display the mean R-squared from regressions including *département*, municipality, and Large Neighborhood fixed effects, respectively.

Table 6: Determinants of victimization: including large neighborhood fixed effects

	Household victimization							Individual victimization			
	Burglary (1)	Car Theft (2)	Motorbike Theft (3)	Bike Theft (4)	Home Vandalism (5)	Car Vandalism (6)	Theft of car objects (7)	Robbery (8)	Theft (9)	Assault (10)	
Share of Immigrants	-0.118** (0.043)	0.024 (0.038)	0.049 (0.165)	0.079 (0.056)	0.024 (0.041)	-0.038 (0.063)	0.036 (0.035)	-0.022* (0.013)	0.017 (0.021)	0.006 (0.020)	
Median Income (log)	-0.026* (0.014)	-0.001 (0.012)	0.074 (0.051)	0.024 (0.018)	-0.012 (0.013)	-0.006 (0.020)	0.009 (0.011)	-0.004 (0.004)	0.004 (0.007)	0.011* (0.006)	
Unemployment rate	0.105** (0.044)	0.007 (0.039)	0.181 (0.169)	0.059 (0.056)	-0.003 (0.042)	0.216*** (0.064)	-0.017 (0.035)	0.026* (0.013)	-0.026 (0.022)	0.041** (0.020)	
Share Monoparental	-0.027 (0.035)	0.061** (0.030)	-0.032 (0.126)	0.070 (0.043)	-0.018 (0.034)	0.152** (0.050)	0.048* (0.027)	0.009 (0.011)	0.016 (0.017)	0.021 (0.016)	
Share Public Housing	-0.008 (0.012)	-0.024** (0.011)	0.071 (0.047)	-0.021 (0.016)	0.028** (0.012)	-0.040** (0.018)	-0.004 (0.010)	-0.007* (0.004)	0.001 (0.006)	0.004 (0.006)	
Share Recent Movers	0.042 (0.031)	0.021 (0.026)	0.021 (0.115)	0.060 (0.038)	0.089** (0.030)	0.104** (0.044)	0.050** (0.024)	0.014 (0.010)	0.036** (0.015)	0.003 (0.014)	
Share 14-18 y.o.	0.098 (0.096)	0.087 (0.082)	-0.099 (0.339)	0.040 (0.115)	0.095 (0.092)	0.141 (0.135)	0.076 (0.074)	-0.027 (0.029)	-0.060 (0.047)	-0.038 (0.044)	
N	63,586	52,453	7,891	35,665	65,588	52,428	52,455	63,655	63,653	63,649	

In addition to the IRIS characteristics reported here, the regressions include the following controls: city density (log), type of neighborhood, household characteristics, household head (respectively individual characteristics) in the household (respectively individual) victimization regressions. They also control for year and large neighborhood fixed effects.  
\* p<0.10, \*\* p<0.05, \*\*\* p<0.001

Table 7: Correlation between IRIS characteristics

	Share Immigrants	Median Income	Unemployment Rate	Share Single Parent	Share Public Housing	Share Recent Movers	Share 14-18 y.o.
Share Immigrants	1.000						
Median Income	-0.255	1.000					
Unemployment Rate	0.528	-0.645	1.000				
Share Single Parent	0.496	-0.457	0.671	1.000			
Share Public Housing	0.505	-0.527	0.661	0.697	1.000		
Share Recent Movers	0.066	0.049	0.126	0.191	-0.104	1.000	
Share 14-18	0.128	-0.158	0.225	0.161	0.292	-0.151	1.000

These correlations are obtained using one observation per IRIS per year. The numbers in the columns of the first line correspond to the numbers in the lines of the first column. For instance, (1) stands for the Share of Immigrants, so that "-2.255" is the correlation between the share of immigrants in an IRIS in a given year and the median income in the same IRIS and year.

Table 8: Unemployment rate in adjacent neighborhoods.

	Household victimization					Individual victimization				
	Burglary Theft (1)	Car Theft (2)	Motorbike Theft (3)	Bike Theft (4)	Home Vandalism (5)	Car Vandalism (6)	Theft of car objects (7)	Robbery (8)	Theft (9)	Assault (10)
<b>No other controls</b>										
$u_{i,IRIS}$	0.005 (0.025)	0.049** (0.024)	0.273** (0.115)	0.116** (0.038)	0.096*** (0.026)	0.353*** (0.042)	0.055* (0.023)	0.016* (0.009)	-0.012 (0.013)	0.051*** (0.013)
$u_{i,IRIS1}$	0.104* (0.058)	-0.009 (0.054)	0.145 (0.266)	0.054 (0.086)	0.034 (0.059)	0.133 (0.095)	0.286*** (0.052)	0.016 (0.020)	0.041 (0.031)	-0.022 (0.029)
$u_{i,IRIS2}$	-0.028 (0.089)	0.049 (0.082)	-0.372 (0.362)	-0.093 (0.127)	-0.055 (0.091)	-0.211 (0.143)	0.140* (0.079)	-0.005 (0.031)	0.014 (0.047)	-0.035 (0.044)
<b>Full specification</b>										
$u_{i,IRIS}$	0.081** (0.028)	0.021 (0.027)	0.211 (0.128)	0.021 (0.042)	0.124*** (0.029)	0.235*** (0.047)	-0.006 (0.026)	0.013 (0.010)	-0.018 (0.015)	0.029** (0.014)
$u_{i,IRIS1}$	0.144** (0.062)	0.035 (0.058)	-0.045 (0.283)	0.030 (0.090)	0.024 (0.063)	0.112 (0.100)	0.226*** (0.055)	0.018 (0.022)	0.051 (0.032)	-0.011 (0.031)
$u_{i,IRIS2}$	-0.013 (0.094)	0.052 (0.087)	-0.475 (0.379)	0.166 (0.132)	0.063 (0.096)	-0.189 (0.150)	0.001 (0.083)	0.004 (0.033)	0.028 (0.049)	-0.052 (0.047)
N	41,781	32,191	4,264	20,502	41,780	32,170	32,194	41,826	41,826	41,822

In these regressions, the sample has been restricted to observations living in actual IRISes, so that the remaining IRISes are homogeneous in terms of size. " $u_{i,IRIS}$ " is the local (IRIS) unemployment rate, " $u_{i,IRIS1}$ " the average unemployment rate in the first ring of adjacent IRISes, and " $u_{i,IRIS2}$ " the average unemployment rate in the second ring of adjacent IRISes. In the second set of results, the controls are environment characteristics (city density (log), type of neighborhood, size of the urban unit), year fixed effects, household characteristics (income, ownership status, number of children) and individual characteristics, corresponding to the surveyed individual for individual victimization and to the household head for household victimization (nationality, age (log), gender, employment status, socio-economic category). They also include large neighborhood fixed effects. Standard errors clustered at the IRIS level reported in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

Table 9: Unemployment rate in adjacent neighborhoods, excluding Paris, Lyon and Marseille

	Household victimization						Individual victimization			
	Burglary Theft	Car Theft	Motorbike Theft	Bike Theft	Home Vandalism	Car Vandalism	Theft of car objects	Robbery	Theft	Assault
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$u_{iris}$	0.074** (0.029)	0.042 (0.028)	0.243* (0.129)	0.003 (0.042)	0.126*** (0.030)	0.227*** (0.048)	0.002 (0.026)	0.017* (0.010)	-0.014 (0.014)	0.026* (0.015)
$u_{iris}$ 1	0.164** (0.064)	0.057 (0.059)	0.031 (0.288)	0.008 (0.090)	0.053 (0.066)	0.104 (0.103)	0.213*** (0.056)	0.017 (0.021)	0.037 (0.031)	-0.009 (0.032)
$u_{iris}$ 2	-0.033 (0.097)	0.073 (0.088)	-0.308 (0.377)	-0.216 (0.132)	-0.076 (0.100)	-0.199 (0.153)	0.011 (0.083)	0.023 (0.032)	0.025 (0.047)	-0.052 (0.049)
N	38,409	30,409	3,945	19,525	38,405	30,390	30,412	38,451	38,450	38,448

In these regressions, the sample has been restricted to observations living in actual IRISes, so that the remaining IRISes are homogeneous in terms of size. " $u_{iris}$ " is the local (IRIS) unemployment rate, " $u_{iris}$  1" the average unemployment rate in the first ring of adjacent IRISes, and " $u_{iris}$  2" the average unemployment rate in the second ring of adjacent IRISes. The controls are environment characteristics (city density (log), type of neighborhood, size of the urban unit), year fixed effects, household characteristics (income, ownership status, number of children) and individual characteristics, corresponding to the surveyed individual for individual victimization and to the household head for household victimization (nationality, age (log), gender, employment status, socio-economic category). They also include large neighborhood fixed effects. \* p<0.10, \*\* p<0.05, \*\*\* p<0.001

Table 10: Largest unemployment rate in adjacent neighborhoods.

		Household victimization				Individual victimization			
Burglary		Car Theft	Home Vandalism	Car Vandalism	Theft of car objects	Robbery	Theft	Assault	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)		
<b>Panel A: Main sample</b>									
$u_{iris}$	0.079** (0.026)	0.026 (0.026)	0.099*** (0.030)	0.205*** (0.046)	-0.005 (0.024)	0.012 (0.010)	-0.015 (0.012)	0.027* (0.014)	
$Max(u_{iris1})$	0.033 (0.022)	0.015 (0.021)	-0.039* (0.021)	-0.082*** (0.033)	0.010 (0.020)	-0.002 (0.007)	0.028** (0.011)	0.006 (0.010)	
$Max(u_{iris2})$	-0.013 (0.016)	0.047** (0.017)	-0.001 (0.018)	-0.026 (0.028)	-0.008 (0.017)	0.008 (0.007)	0.012 (0.010)	-0.012 (0.008)	
<b>Panel B: Excluding Paris, Lyon and Marseille</b>									
$u_{iris}$	0.075** (0.028)	0.046* (0.025)	0.101** (0.031)	0.194*** (0.048)	-0.005 (0.025)	0.017* (0.010)	-0.011 (0.012)	0.023 (0.015)	
$Max(u_{iris1})$	0.048** (0.024)	0.024 (0.021)	-0.044** (0.022)	-0.075** (0.034)	0.005 (0.021)	0.004 (0.007)	0.021* (0.011)	0.006 (0.011)	
$Max(u_{iris2})$	-0.002 (0.017)	0.051** (0.018)	-0.006 (0.019)	-0.023 (0.030)	-0.013 (0.019)	0.001 (0.005)	0.017* (0.010)	-0.012 (0.009)	

In these regressions, the sample has been restricted to observations living in actual IRISes, so that the remaining IRISes are homogeneous in terms of size. " $u_{iris}$ " is the local (IRIS) unemployment rate, " $Max(u_{iris1})$ " the largest unemployment rate in the first ring of adjacent IRISes, and " $Max(u_{iris2})$ " the largest unemployment rate in the second ring of adjacent IRISes. In both panels, the controls are environment characteristics (city density (log), type of neighborhood, size of the urban unit), year fixed effects, household characteristics (income, ownership status, number of children) and individual characteristics, corresponding to the surveyed individual for individual victimization and to the household head for household victimization (nationality, age (log), gender, employment status, socio-economic category). They also include large neighborhood fixed effects. The second panel is obtained after excluding Paris, Lyon and Marseille from the sample. Standard errors clustered at the IRIS level reported in parentheses. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.001$

## References

- Bayer, P., Ross, S. L., and Topa, G. (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy*, 116(6):1150–1196.
- Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76:169–217.
- Bell, B., Machin, S., and Fasani, F. (2010). Crime and immigration: Evidence from large immigrant waves. IZA Discussion Papers 4996, Institute for the Study of Labor (IZA).
- Bernasco, W. and Luykx, F. (2003). Effects of attractiveness, opportunity and accessibility to burglars on residential burglary rates of urban neighborhoods. *Criminology*, 41(3):981–1002.
- Bianchi, M., Buonanno, P., and Pinotti, P. (2012). Do immigrants cause crime? *Journal of the European Economic Association*, 10(6):1318–1347.
- Bruinsma, G. J. N. (2007). Urbanization and urban crime: Dutch geographical and environmental research. *Crime and Justice*, 35(1):453–502.
- Buonanno, P., Montolio, D., and Vanin, P. (2009). Does social capital reduce crime? *Journal of Law and Economics*, 52(1):145–170.
- Calvó-Armengol, A., Verdier, T., and Zenou, Y. (2007). Strong and weak ties in employment and crime. *Journal of Public Economics*, 91(1-2):203–233.
- Cohen, L. E. and Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44(4):588–608.
- Ehrlich, I. (1973). Participation in illegitimate activities: A theoretical and empirical investigation. *Journal of Political Economy*, 81(3):521–65.
- Evans, W. N., Oates, W. E., and Schwab, R. M. (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy*, 100(5):966–91.
- Fougère, D., Kramarz, F., and Pouget, J. (2009). Youth unemployment and crime in France. *Journal of the European Economic Association*, 7(5):909–938.
- Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996). Crime and social interactions. *The Quarterly Journal of Economics*, 111(2):507–48.
- Gould, E. D., Weinberg, B. A., and Mustard, D. B. (2002). Crime rates and local labor market opportunities in the United States: 1979-1997. *The Review of Economics and Statistics*, 84(1):45–61.
- Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. *Journal of International Economics*, 67(1):1 – 24.
- Harris, C. D. (1954). The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers*, 44(4):315–348.
- Hémet, C. and Malgouyres, C. (2018). Diversity and Employment Prospects: Neighbors Matter! *Journal of Human Resources*, 53(3):825–858.

- Kelly, M. (2000). Inequality and crime. *The Review of Economics and Statistics*, 82(4):530–539.
- Kling, J. R., Ludwig, J., and Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *The Quarterly Journal of Economics*, 120(1):87–130.
- Ludwig, J., Duncan, G. J., and Hirschfield, P. (2001). Urban poverty and juvenile crime: Evidence from a randomized housing-mobility experiment. *The Quarterly Journal of Economics*, 116(2):655–679.
- Machin, S. and Meghir, C. (2004). Crime and economic incentives. *Journal of Human Resources*, 39(4).
- Patacchini, E. and Zenou, Y. (2008). The strength of weak ties in crime. *European Economic Review*, 52(2):209–236.
- Shaw, C. and McKay, H. (1942). *Juvenile Delinquency and Urban Areas*. University of Chicago Press.
- Solignac, M. and Tô, M. (2018). Do workers make good neighbours? the impact of local employment on young male and female entrants to the labour market. *Annals of Economics and Statistics*, (130):167–198.
- Spenkuch, J. L. (2010). Understanding the impact of immigration on crime. MPRA Paper 22864, University Library of Munich, Germany.
- Zenou, Y. (2005). Crime, location and the housing market. CEPR Discussion Paper 5389, Center for Economic Policy Research.

**D.6 Custodial vs. Non-Custodial Punishments: Measuring the Impact of an Anticipated Reform**

Custodial vs. Non-Custodial  
Punishments:  
Measuring the Impact of an Anticipated Reform

Bastien Michel  
Paris School of Economics  
[bastien.michel@psemail.eu](mailto:bastien.michel@psemail.eu)

Camille Hémet  
Paris School of Economics  
[camille.hemet@psemail.eu](mailto:camille.hemet@psemail.eu)

Work in progress: please do not circulate.

October 2019

**Abstract**

We study a reform of the Danish legislation after which most offenders tried for drunk-driving received a non-custodial sanction instead of a custodial one. We show that stakeholders anticipated the consequences of the reform as selection is observed in the nature of the cases tried before and after the reform. To measure its impact, we use an instrumental variable approach exploiting the fact that the closer to the reform a crime was committed, the more likely the defendant was to be tried after. Offenders who received a non-custodial sentence commit fewer crimes and have stronger ties to the labor market.

**JEL Codes:** K14, K4

---

Acknowledgements: we would like to thank Timo Hener, Randi Hjalmarsson, Nicolai Kristensen, Elena Mattana, Anna Piil Damm, Victor Ronda, Michael Rosholm, and Marianne Simonsen for their useful comments and suggestions, as well as the District Courts in Aarhus and Odense and the Danish Prison and Probation Service for supplying information about relevant institutional details. Funding from TrygFonden's Centre for Child Research is greatly appreciated. We would also like to thank seminar participants at Aarhus University, Aix-Marseille School of Economics, and the Rockwool Foundation for their useful comments. The usual disclaimer applies.

## I. Introduction

Every year, 1.35 million people die and 20 to 50 million more suffer non-fatal injuries as a result of a road traffic accident (WHO, 2018). Driving under the influence of alcohol (referred to as *drunk driving* hereafter) is a major risk factor for road traffic accidents and, as such, represents an important public health issue in most countries. For instance, throughout the world, drunk driving is believed to account for more than 273,000 deaths every year (Vissers, 2017).<sup>1,2</sup> In the European Union and in the United States, alcohol would have caused 25 to 30% of all road fatalities in 2015 (European Commission, 2015; NHTSA, 2017) – representing around 6,400 and 10,265 fatalities respectively.<sup>3</sup> A key challenge to prevent drunk driving lies in the addiction problems drunk drivers are particularly prone to<sup>7</sup> and, as a consequence, what may work to prevent other crimes may not apply for drunk drivers. While prevention policies have managed to reduce the number of road traffic accidents involving alcohol in many countries over the past 50 years, evidence still remains extremely scarce on how best to sanction drunk drivers in order to prevent recidivism.

Our paper fills this gap by providing first-time robust evidence on the relative impact of a custodial sentences (e.g. incarceration) or non-custodial ones (e.g. probation) on drink-drivers.<sup>10</sup> In order to do so, we study a large-scale reform of the Danish legislation implemented in 2000, whereby jail time (a custodial sentence) was replaced by a probation period (a non-custodial sentence) for most drunk-driving crimes.<sup>12</sup> We measure the impact of the reform using rich administrative datasets and find that substituting a non-custodial sentence for a custodial one does not have any impact on offenders' probability of committing another drunk-driving crime or on the number of drunk-driving crimes they commit. We also find that incarceration increases the average number of other crimes drunk-driving offenders subsequently commit, which suggests that the criminogenic effects of custodial sentences were found to have on other offenders (Cullen et al., 2011; Aizer and Doyle, 2015) impinge on a broad set of offenders, including subsets who exhibit relatively low proclivity for criminal behaviors

---

<sup>1</sup> The share of alcohol-related road fatalities for different countries is estimated to range between 5% and 35% (Vissers, 2017).

<sup>2</sup> In consequence, drunk driving also represents a significant cost for most countries. For instance, in the United States, the economic cost of all alcohol-impaired accidents was estimated at 44 billion dollars for the sole year of 2010 (NHTSA, 2017).

<sup>3</sup> Calinescu and Admainite (2018) estimated that 80% of these deaths could have been avoided had all the drivers been sober.

<sup>7</sup> Evidence on the actual share of convicted drink-drivers suffering from an alcohol abuse problem is not very conclusive but suggests that it is important (for instance, see Vingilis (1983), Lapham et al. (2001), and Chou et al. (2006)). In the context of Denmark, a study revealed that two thirds of all drunk-driving offenders were suffering from an alcohol abuse problem in 1990 (Kramp et al., 1990).

<sup>10</sup> The Oxford English dictionary defines a custodial sentence as “a judicial sentence requiring an offender to be imprisoned, as opposed to a fine, a community service order, etc.” (accessed in March 2019).

<sup>12</sup> As further discussed below, the probation period was associated with either a mandatory participation in a rehabilitation program or some community service (depending on whether or not the offender suffered from alcohol addiction).

such as drunk drivers. Finally, we find that, compared to non-custodial sentences, custodial sentences significantly weaken offenders' labor market attachment as they reduce both their annual number of days worked and earnings. This further documents the relative effects of custodial and non-custodial sentences on key dimensions of offenders' lives (Western et al., 2001), and shed some light on a possibly important mechanism for offenders' subsequent involvement in criminal activities.

These findings contribute to the broad literature on the relative impact of custodial and non-custodial sentences (Chalfin and McCrary, 2017) and to the more limited one focusing specifically on drunk drivers. To our knowledge, the only other study focusing on the impact of sanctions on drunk drivers is Hansen (2015), who exploited discontinuities in the severity of the punishment based on administrative blood alcohol content thresholds in the State of Washington (USA) to measure the impact of incarceration and incarceration length on the recidivism rate. He found lower rates of recidivism among drivers liable to incarceration (and more severe sanctions if re-arrested for drunk driving) than among drivers not liable to any legal sanction, as well as among drivers liable to a longer period of incarceration than among drivers liable to shorter ones.<sup>13</sup> Although important, Hansen's findings say nothing about the relative impact of different *types* of sanctions on drunk drivers and are of little help to policymakers wondering whether drunk drivers should receive a custodial or a non-custodial sentence.

As a second contribution, we show that individuals and, in particular, the wealthiest ones can “game the system” in order to benefit from more favorable circumstances around the time of a policy change. To our knowledge, we are the first to provide such an evidence in the context of a justice reform. Indeed, we find evidence that, in the context of the studied reform, offenders anticipated that they could avoid prison by postponing their trial until after the reform (a feature of Danish legislation guarantees that defendants tried after a reform for a crime committed prior to it must be tried under the more lenient of the two laws). In practice, these anticipations materialized through a sharp drop in the number of cases tried from the moment the law was signed (but before it actually entered into force), as well as a linear decrease in the share of defendants receiving a custodial sentence. Moreover, we also show that wealthier defendants were more likely to have their trial delayed and to avoid prison. These findings suggest that traditional quasi-experimental estimators should be used with caution in similar contexts where salient contextual changes (such as a legislative reform, a program

---

<sup>13</sup> Investigating possible explanatory mechanisms, the author concluded that these effects could not be attributed to either an incapacitation effect or a rehabilitation one but rather to a specific deterrence effect. This suggests that unconditional prison sentences are effective at deterring recidivism.

scale-up, etc.) can be anticipated by their stakeholders and that, in such a context. Incidentally, our findings also question the degree of fairness with which cases can be handled in times of legislative changes, thus adding to the growing literature documenting sources of dysfunction in justice systems (Vidmar, 2011; Danziger et al., 2011; Abrams et al., 2012; Anwar et al., 2012; Anwar et al., 2014; Philippe and Ouss, 2018; Cohen and Yang, 2019).

In order to bypass this selection problem and identify the relative impact of custodial and non-custodial sentences, we use a novel instrumental variable: the time elapsed between the date the crime was committed and the date the reform entered into force. The idea behind this instrument is that the closer to the reform a crime was committed, the more likely the defendant was to be tried under the new law and placed on probation instead of being incarcerated. We document that there is indeed significant variation in the probability of being tried after the reform among offenders arrested up to one year before the reform. In addition, the variation in this variable entirely relies on the date of occurrence of the crime, which we argue to be plausibly exogenous.<sup>17</sup> In particular, we show that defendants' characteristics are not correlated with the instrument, suggesting that the independence assumption is likely to be met. Moreover, in order for the exclusion restriction to be satisfied, the *timing* of the crime itself must not have any direct impact on defendants' subsequent crime and labor market outcomes. In particular, one could be concerned that alleged drunk-drivers do not share the same characteristics depending on the period when the crime was committed. To mitigate this concern, we control for trend and seasonality effects by including individuals tried between 24 and 13 months prior to the reform in our sample.

The rest of the paper is organized as follows: in section II, we provide information on the reform and the context in which it was implemented; in section III, we provide a description of the data we use; in section IV, we describe the implementation of the reform; in section V, we describe our identification strategy; in section VI, we present our results; finally, section VII is the conclusion.

---

<sup>17</sup> An approach loosely related to the one used in Drago et al. (2009). In their study, they used the Collective Clemency Bill passed by the Italian Parliament in July 2006 to measure the impact of suspended sentence length on recidivism. This reform reduced the length of the prison sentence of all inmates who had committed a crime before May 2, 2006. As a consequence, about 40 percent of the prison population of Italy were released from prison on August 1, 2006 under the condition that they would have to serve the remaining of their sentence if they were to commit another crime in the 5-year period following their release. Interestingly, the length of offenders' suspended sentence varied depending on their prison entry date, which the researchers argued is exogenous and which they used to measure the impact of suspended sentence length on recidivism.

## II. Background

### *II.A Drunk-driving legislation (until 2000)*

In the last quarter of the 20<sup>th</sup> century, legislations on drunk-driving crimes were gradually hardened throughout the world. In many countries, this evolution resulted in the introduction of an administrative blood alcohol content threshold and of a legislation specifying the legal sanctions incurred by drink drivers depending on the severity of their offense. In turn, these reforms significantly increased the number of individuals sent to prison for a drunk-driving crime, as well as the associated costs for the tax-payers.

In Denmark, a first blood alcohol content threshold was established at 0.08 mg/l in 1976.<sup>19</sup> In 1998, it was lowered to 0.05 mg/l. Following that reform, all individuals facing a prison sentence were tried in a court of justice (as is still the case) and almost all individuals tried for an alleged drunk-driving crime were tried in first instance in one of Denmark's 84 district courts.<sup>20</sup> As displayed in *Table 1*, very few of the defendants tried for a drunk-driving crime avoided a conviction. Indeed, only 1.3% of the defendants tried in 1999 managed to avoid all of the legal sanctions displayed in the table. Among those who were sanctioned, the severity of the punishment varied quite substantially depending on the characteristics of the offense (driver's level of impairment at the moment of the crime and whether or not aggravating circumstances could be found<sup>21</sup>) and the number of prior drunk-driving convictions. Still, most of the individuals tried for an alleged drunk-driving crime received an unconditional prison sentence (71.9%).<sup>22</sup> In 95.3% of these cases, the length of the unconditional prison sentence remained below 60 days.<sup>23</sup> Fines and suspensions of the driving license were also frequently used to sanction drink-drivers, as they were requested in 27.8% and 30.8% of the cases

---

<sup>19</sup> For contextual purposes, it is worth noting that in Denmark, the minimum drinking age and minimum age to purchase alcoholic beverages are both 18, which is also the minimum licensing age for drivers of private vehicles.

<sup>20</sup> At that time, the vast majority of criminal cases were tried in a district court in first instance. The only few drunk-driving cases which were tried in one of the two High Courts in first instance were those for which a prosecutor had requested more than four years of imprisonment. These very rare cases were not affected by the reform studied in this article and therefore fall out of the scope of this study.

<sup>21</sup> For instance, aggravating circumstances were found when the driver had previously had their driving license suspended, caused personal injury, damaged property, or exposed a person or their property to such injuries or damages through gross disregard of road safety.

<sup>22</sup> For instance, the sanction for a first drunk-driving crime could range from a simple fine for offenders with a BAC between 0.05 and 0.08 mg/l to imprisonment for offenders with a BAC over 0.20 mg/l, along with a suspension or revocation of their driving license. Incarceration could also be requested against drivers with lower BAC levels if the offense constituted a repeat offense and/or if aggravating circumstances could be found.

<sup>23</sup> The length of the prison sentence only exceeded 60 days for very severe cases, as displayed in *Table A.1* (placed in the appendix), which describes the sanctions incurred by a defendant based on the nature of their crime.

respectively. Conversely, conditional prison sentences (probation) and community work were either rarely used or not used at all to sanction drink-drivers.

Offenders who received an unconditional prison sentence usually served their time in one of the country's open prisons<sup>24</sup> – low-security prisons where fences, walls, and barriers are minimal, offering softer detention conditions for the inmates.<sup>25</sup> However, offenders suffering from an alcohol abuse problem who received a prison sentence of no more than 60 days could ask to benefit from a *pardon scheme* as part of which their unconditional prison sentence could be commuted to a two-year probation period and a mandatory participation in a yearlong rehabilitation program (described below).<sup>26,27</sup>

## *II.B The 2000 reform*

In 2000, a reform was passed introducing cheaper, more lenient sentences against drink-drivers.<sup>28</sup> As part of it, custodial sentences of no more than 60 days were replaced by non-custodial ones – a two-year probation period and a fine, combined with either mandatory participation in a yearlong rehabilitation program (identical in every single way to the one offered as part of the pardon scheme just mentioned above) or community service.<sup>29</sup>

The choice of mandatory participation in a rehabilitation program or community service was left to the judges based on whether or not the offender suffered from an alcohol abuse problem, with participation in the rehabilitation program being reserved for offenders exhibiting an alcohol abuse

---

<sup>24</sup> Individuals serving a prison sentence inferior to four years and presenting no security threats to prison staff and other inmates are usually incarcerated in open prisons.

<sup>25</sup> In practice, prisoners can apply for early release on parole after having served two-thirds of their sentence (as long as more than two months have been served), and most of them are released at that time (Kyvsgaard, 2004). Under exceptional circumstances (such as health reasons, specific conditions in inmates' family, etc.), prisoners may be released after having served half of their sentence but, again, as long as they have served more than two months.

<sup>26</sup> Indeed, from 1990, drunk-driving offenders with "a strong need for rehabilitation" who were sentenced to no more than 40 days of incarceration could apply to the Danish Prison and Probation Service to see their unconditional prison sentence commuted into a two-year probation period, starting with mandatory participation in a yearlong rehabilitation program. The sanction came together with a fine and could also be combined with either a suspension or a revocation of offender's driving license. In 1994, the pardon scheme was extended to drunk-driving offenders sentenced to no more than 60 days of incarceration but remained applicable to offenders exhibiting a "strong need for rehabilitation" only. Towards the end of the 1990s, the share of offenders enrolled in the program who were granted a pardon upon successful completion of the program remained between 75 and 80%.

<sup>27</sup> According to the Prison and Probation service, only around 750 offenders were pardoned each year under the pardon scheme (Kriminalforsorgens årsberetning, 1998 and 1999), suggesting that around 70% of drunk-driving offenders who received an unconditional prison sentence were incarcerated. Using a different data source, Clausen (2007) estimated that 58.2% of all drunk-driving offenders sentenced to no more than 60 days did not benefit from the pardon scheme in the 18-month period preceding the end of the reform. This lower rate does not change the conclusion that a large share of offenders did not benefit from the pardon scheme prior to the change in the legislation in 2000.

<sup>28</sup> Nielsen and Kyvsgaard (2007) reported that, in 2004, the total cost associated with the alternative sanction averaged at 8.300 DKK per offender (including the costs associated with the offenders' supervision and rehabilitation program), while the cost associated with the prison sentence they would have served had they been convicted prior to the reform averaged at 15.800 DKK per offender.

<sup>29</sup> More information on the reform can be found on the following webpage (in Danish): <https://www.retsinformation.dk/Forms/R0710.aspx?id=88116> (accessed in March 2019)

problem. As part of this program, offenders had to take a drug causing an acute sensitivity to ethanol and to participate in an alcohol treatment program.<sup>30</sup> Offenders were monitored throughout the duration of the treatment and the following one-year probation period.<sup>31</sup> Probation officers were in charge of verifying that the terms of the probation were being respected and, in particular, of controlling offenders' drug intake and participation in the alcohol treatment program during the first phase of the scheme. Community service was to be requested against offenders who did not exhibit such an alcohol problem and was substituted to the former sentences at the following rate: 30 hours for 10 to 14 days of imprisonment, 40 hours for 20-30 day sentences, and 60 hours for 40 to 50 days in jail.

It is important to note that this reform seems to have been perceived as a softening of the legislation by the vast majority of offenders – as intended by the administration. Indeed, while the incarceration conditions in Scandinavian prisons are considered to be quite exceptional by American and European standards (Lappi-Seppälä, 2007; Pratt, 2008; Pratt and Eriksson, 2011; Ward et al., 2013), inmates remain subject to important freedom restrictions and other usual discomforts associated with imprisonment, even when they are incarcerated in an open prison. Moreover, while the length of the probation period imposed on offenders after the reform is substantial in comparison to the rather short duration of the incarceration spell faced by offenders before the reform, the share of offenders who broke the terms of their parole remained limited, suggesting that few offenders perceived incarceration as preferable to the alternative sanction. Only 10.1% of the offenders who were placed on parole in the 6-month period following the reform were incarcerated at least once in the 2-year period following their trial.

However, the reform left unchanged the punishments incurred by offenders facing no prison sentence, as well as by those facing more than 60 days of imprisonment, who kept on serving their prison sentence after the reform (mostly offenders facing extreme aggravating circumstances, and extreme repeat drink-drivers). Moreover, the reform did not apply systematically to offenders who had already been placed on probation for a drunk-driving crime more than once or to those who were on probation at the time of the crime for an alcohol-related crime.

---

<sup>30</sup> In practice, this program could take a variety of forms (ranging from group sessions at a clinic to individual meetings with general practitioners) and could vary in intensity depending on individuals' location, needs, and motivation (Nielsen and Kyvsgaard, 2007).

<sup>31</sup> During the first two months of the two-year program, offenders would usually meet with their probation officers every 2 weeks, but would only meet once a month thereafter – unless arguments in favor of a more intensive monitoring prevailed.

Finally, as is often the case for important reforms that require a certain level of preparation, a few months elapsed between the moment the law was signed and the moment it entered into force: while the law was signed by the Parliament on April 4<sup>th</sup>, 2000, it only entered into force on July 1<sup>st</sup>, 2000 (referred to as *the date of the reform* hereafter).<sup>32</sup>

### III. Data<sup>33</sup>

In order to document the impact of the reform, we use Denmark's rich administrative datasets, which contain individual-level information on all of the country's residents since 1980. These datasets include a wide range of information collected annually, which can be merged using unique individual identifiers. We use these datasets to identify alleged drunk-driving crimes committed and tried around the time of the reform, to compute our outcome variables, and to create the set of control variables we use as covariates.

#### III.A Administrative datasets

##### *Information on crime, charges, and sanctions*

Danish administrative data include detailed information on individuals' involvement in criminal activities since 1981. In particular, they include information on all crimes reported to the police, including information about the identity of the alleged perpetrators, as well as the date and nature of the *main* crime (in cases where several crimes were reported concomitantly against a single person). They also include information on all charges pressed by a prosecutor, including the identity of the individuals against which the charges were pressed, and the date and nature of the *main* charge (in cases where several charges were pressed concomitantly against a single person).<sup>34</sup> Finally, they also contain information about the outcome of every criminal case settled by the police, a prosecutor, or a judge. For each decision of justice, information is available on the identity of the defendants, the entity responsible for issuing the sanction, the date when the sanction was issued, as well as the nature

---

<sup>32</sup> For instance, the 1998 law on drunk-driving crimes mentioned earlier was signed on June 10<sup>th</sup>, 1997 but only entered into force on March 1<sup>st</sup>, 1998.

<sup>33</sup> The administrative registers used as part of this project are the following ones: BEF, FAM, IDAN, IEPE, INDH, KRAN, KRIN, KRSI, and UDDA. Descriptions of the different registers can be found on the following webpage: [http://www.dst.dk/da/TilSalg/Forskningsservice/Data/Register\\_Variabeloversigter](http://www.dst.dk/da/TilSalg/Forskningsservice/Data/Register_Variabeloversigter) (accessed in March 2019). We provide further information on each of these datasets in *Table A.2* (placed in the appendix).

<sup>34</sup> These crime and charge codes are recorded by the police using a detailed 7-digit hierarchical code (1,161 codes) – the last three digits often indicating the severity of the crime. For instance, 60 different codes can be used to categorize drunk-driving crimes and charges (29 of which were effectively encountered during the study period).

of the sanction (a prison sentence, a fine, a withdrawal, an acquittal, etc.) and its severity (fine amounts, conditional and unconditional prison sentence length, etc.).

Unfortunately, whether or not an offender was actually incarcerated and the length of their incarceration spell is only imprecisely observed. That is why we focus on whether or not an individual *received* a custodial sentence in what follows – and not on whether or not they were actually incarcerated. However, we construct a proxy for whether or not an individual was actually incarcerated and obtain results similar to those displayed below when using it as our variable of interest.<sup>35,36</sup> Finally, drivers' blood alcohol content at the time of their arrest is not available in the datasets.

#### *Information on labor market attachment*

Danish administrative data also include information on all residents' labor market attachment, which is measured and collected every year in November by Statistics Denmark.

### III.B Outcome variables

Using these registers, we compute the two groups of outcome variables on which we focus as part of this study: crime-related and labor-related outcomes.

First, we use these administrative records to compute two outcomes indicative of offenders' post-sentencing criminal activity. We start by assessing the relative impact of custodial and non-custodial sentences on offenders' involvement in subsequent drunk-driving crimes. In order to do so, we calculate whether or not individuals were convicted again of another drunk-driving crime and, if they did, the number of such crimes they committed. Then, in order to investigate any criminogenic effect of incarceration, we also compute an outcome indicating whether or not individuals were convicted of any other crimes and, if they did, the number of such crimes they committed. We measure the impact of the reform on these outcomes from 3 months to 10 years, from the date when the drunk-driving case was settled in court.<sup>37</sup>

---

<sup>35</sup> Our proxy for whether or not an individual was incarcerated captures whether an individual spent at least 10 days in prison – 10 days being the minimum duration of prison sentences requested for a drunk-driving crime, as displayed in *Table A.1* (placed in the appendix). To account for the fact that the length of the incarceration spell is imprecisely estimated, we also consider as having been incarcerated individuals who are observed to have been incarcerated for 9 days.

<sup>36</sup> The only difference is that, logically enough, the results are larger in magnitude. Results are available upon request.

<sup>37</sup> In order to measure the *net* impact of incarceration, we exclude from the calculation of these outcomes any crime registered under the same case ID or related to any other crime committed prior to the decision of justice considered in this study.

Second, we also use these administrative registers to compute outcomes indicative of individuals' attachment to the labor market at different time horizons following the ruling. More specifically, we focus on the relative impact of custodial and non-custodial sentences on the annual number of days worked, as well as on annual earnings before tax and any social contributions (which is inflated to 2015 prices using Statistics Denmark's Consumer Price Index). We measure the impact of the reform on these outcomes at different time horizons, from 1 to 10 years, from the date when the drunk-driving case was settled in court.

### III.C Control variables

Finally, we use these registers to compute the set of control variables that we include in the regressions. This conditioning set provides two types of information on individuals' pre-crime characteristics. First, we use variables indicative of the characteristics of the trial, such as whether the defendant was a juvenile at the time of the crime and the nature of the main charge (using a detailed 7-digit drunk-driving charge code). Second, we also include defendants' background information, such as their gender, age at the time of the trial, immigration status (as per Statistics Denmark's typology: "immigrants", "descendant of immigrants", or "rest of the population"), their past criminal activity (the number of convictions in the 5-year period preceding their crime for other drunk-driving crimes, other road traffic crimes, and non-road traffic crimes), marital status, highest educational achievement, type of job held, and annual earnings (before tax and any social contributions).

Unless specified otherwise, all baseline background characteristics included in the conditioning set were measured at the end of the year preceding the crime and are available for the vast majority of the offenders in our sample.<sup>38</sup>

## IV. Implementation of the reform

In order to determine how to measure the impact of the reform, we start by studying its implementation.

### IV.A Anticipation

In *Figure 1*, we describe how the reform was implemented through the evolution of the following four indicators between 1999 and 2001: a) the number of alleged drunk-driving crimes resulting in a

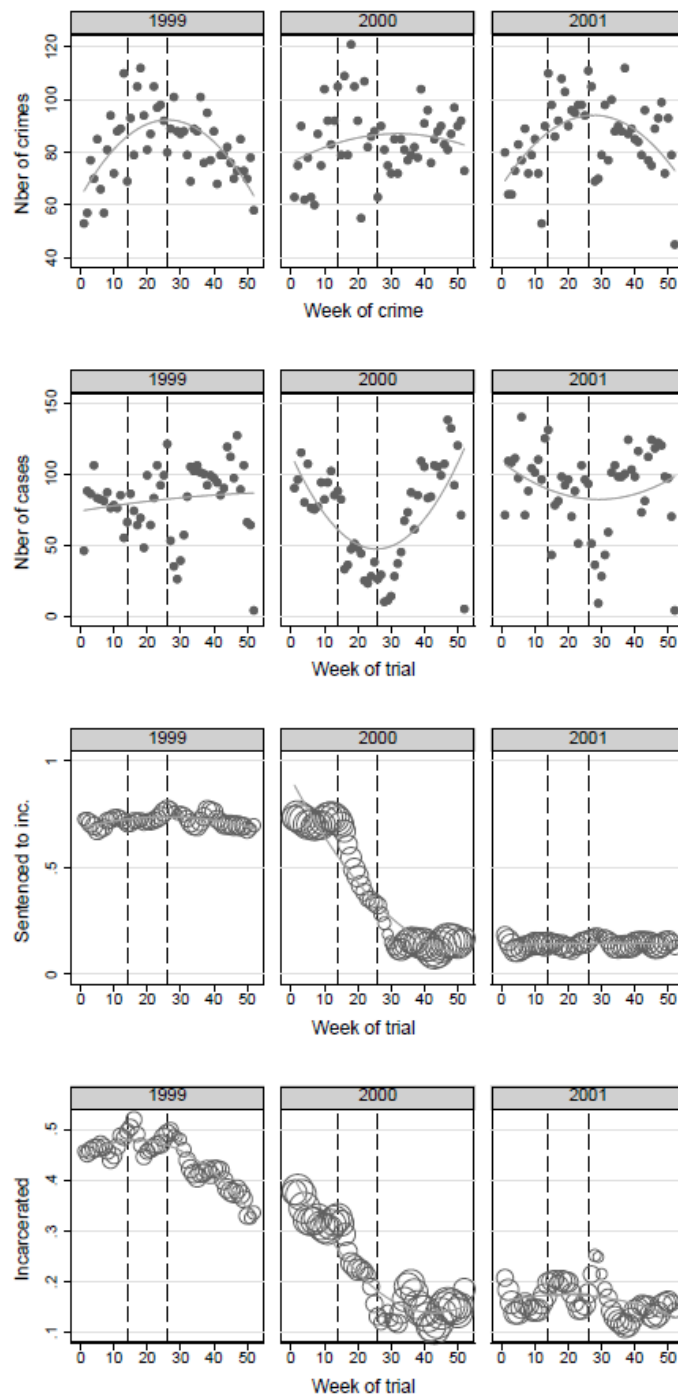
---

<sup>38</sup> The variables included in the conditioning set are all available from 1986.

trial committed every week; b) the number of drunk-driving cases tried every week in district courts; c) the share of defendants tried for drunk-driving who received a custodial sentence by week of trial; d) the share of defendants tried for drunk-driving who were actually incarcerated by week of trial.<sup>39</sup> For each year, the first dotted vertical line marks week 14 (the week when the law was signed in 2000)

---

<sup>39</sup> Again, this variable captures whether an individual has spent at least 10 days in prison – 10 days being the minimum duration of prison sentences requested for a drunk-driving crime. To account for the fact that the length of the incarceration spell is imprecisely estimated, we also consider as having been incarcerated individuals who are observed to have been incarcerated for 9 days.



**Figure 1 – Implementation of the drunk-driving legislation reform**

The consequences of the reform are depicted here through the evolution of the following four indicators around the time of the change in legislation: a) the number of drunk-driving cases tried every week in district courts; b) the number of drunk-driving crimes resulting in a trial committed every week; c) the share of defendants tried for drunk-driving who received a custodial sentence by week of trial; d) the share of defendants tried for drunk-driving who were actually incarcerated by week of trial. For each year, the first dotted vertical line marks the week when the law was signed (week 14) and the second one marks the week when it entered into force (week 26).

and the second one marks week 26 (the week when it entered into force in 2000). The only reform implemented during these three years occurred in 2000.<sup>40</sup>

Strikingly, the evolution of these indicators reveals that the way in which drunk-driving cases were handled in district courts changed drastically in the months preceding the entering into force of the reform. Indeed, the number of cases tried each week dropped significantly from 91.7 cases on average in the three weeks preceding the signing of the law to 28.0 cases on average during the transition period (after the law was signed but before it entered into force) – representing a 69.5% decrease. This is the case despite the fact that there was no such variation in the number of alleged crimes resulting in a trial committed in the preceding months or in the number of cases tried during the same period in adjacent years, 1999 and 2001.<sup>41</sup> This suggests that stakeholders (courts of justice and/or defendants) anticipated the change in legislation and that, as a consequence, a large share of trials were postponed until after the reform. In turn, this implies that a group of offenders who should have been tried before the reform was tried after.<sup>42</sup>

The share of drink-drivers who received a custodial sentence and the share of those who were actually incarcerated also decreased substantially from the moment the bill was signed. Surprisingly, this time, the decline did not take the form of a sharp discontinuity but rather of a linear decrease. Overall, the share of defendants receiving a custodial sentence decreased progressively from around 73.6% on average in the three weeks preceding the signing of the law to 34.5% on average in the three weeks preceding the entering into force of the reform – representing a 53.1% decrease.<sup>43</sup> The share of drink-drivers who were actually incarcerated also decreased substantially from the moment the bill was signed, although from a lower initial level (because of the implementation of the rehabilitation

---

<sup>40</sup> For data confidentiality reasons, indicators c) and d) displayed in *Figure 1* are calculated as moving averages. For each year  $y$ , the value of these indicators (as displayed in *Figure 1*) is calculated as the average value of the indicators over years  $y-1$ ,  $y$ , and  $y+1$ .

<sup>41</sup> The number of cases tried in the week following July 1<sup>st</sup> is low for all three years. This is a result of judges' summer vacation period, during which the number of cases tried in district courts goes down substantially.

<sup>42</sup> In total, had the same number of drunk-driving cases been tried between weeks 14 and 26 in 1999, 2000, and 2001, we estimate that roughly 48.1% of the drunk-driving cases which should have been tried during the transition period were in fact postponed until after the reform. In order to reach this figure, we assume that in the absence of the reform, the number of drunk-driving cases tried in 2000 would have been equal to the average number of such cases tried in the same weeks in 1999 (1,071) and 2001 (1,123) – 1,097. However, only 569 drunk-driving cases were tried during the transition period in 2000, suggesting that around 528 were postponed – which represents 48.1% of what would have been the total number of drunk-driving cases tried during that period.

<sup>43</sup> Had the same number of drunk-driving cases been tried between weeks 14 and 26 in 1999 and 2000, and had the type of crimes tried been similar in the two periods, we estimate that 61.1% of the drunk-driving cases which should have been tried during the transition period and as part of which an unconditional prison sentence would have been rendered were in fact postponed until after the reform. In order to reach this figure, we assume that in the absence of the reform, the number of drunk-driving cases resulting in the offender receiving a prison sentence during the transition period would have been equal to the number of such cases tried during the same period in 1999 (782). Only 303 of the drink-driver cases tried during the transition period received an unconditional prison sentence, which suggests that around 478 offenders who should have been tried during the transition period and should have received an unconditional prison sentence did not – representing 61.1% of what would have been the total number of drink-drivers to receive an unconditional prison sentence during that period.

program described above): the share of defendants who were actually incarcerated before the reform is around a third lower than the share of defendants who received a custodial sentence. Interestingly, while the evolution of this indicator exhibits a pattern similar to the one described above, it seems that the Prison and Probation Service in charge of enforcing the sanctions rendered by the district courts anticipated the entering into force of the reform even further: the share of individuals incarcerated started to decrease a year before the entering into force of the reform.<sup>44</sup>

While we are not able to pin down the exact underlying mechanisms at play here, a closer look at the stakeholders' incentives suggests that both defendants and judges had reasons to postpone drunk-driving cases until after the reform: the former to avoid prison, the latter to reduce the number of cases which might have to be retried. Indeed, an important feature of Danish legislation guarantees that defendants tried after a reform for a crime committed prior to it must be tried under the more lenient of the two laws. In the context of the reform at hand, this means that individuals tried for a crime committed prior to the reform faced the risk of being incarcerated if tried before the reform, while they merely faced the risk of being placed on probation if tried after. The same feature also guarantees that defendants tried prior to the passing of a law lowering the sanction for the crime they were convicted of may request that their case be retried if they are still in prison when the reform enters into force.

#### IV.B Selection

Investigating the characteristics of the defendants who acted in anticipation of the reform and got their case postponed, we find evidence suggesting that the selection was not random. In order to investigate this question, we focus on individuals tried between January 1<sup>st</sup>, 1999 and December 31<sup>st</sup>, 2000, and regress different variables indicative of individuals' criminal priors and labor market attachment on a constant, a dummy variable indicating when a case was tried in 2000, quarter fixed effects, and the interactions between the year dummy and the quarter fixed effects.

---

<sup>44</sup> Again, had the same number of drunk-driving cases been tried between weeks 14 and 26 in 1999 and 2000, and had the type of crimes tried been similar in the two periods, we estimate that 72.1% of the drunk-driving cases which should have been tried during the transition period and as part of which the defendant would have been incarcerated were in fact postponed until after the reform. As a consequence, their defendants avoided prison. In order to reach this figure, we assume that in the absence of the reform, the number of drunk-driving cases resulting in the offender being incarcerated during the transition period would have been equal to the number of such cases tried during the same period in 1999 (513). Only 143 of the drunk-driving cases tried during the transition period led to the actual incarceration of the offender, which suggests that around 370 offenders who should have been tried during the transition period and incarcerated, were not – representing 72.1% of what would have been the total number of drink-drivers actually incarcerated during that period.

In *Table 2*, we report the coefficients associated with the year dummy (which capture changes in the composition of the cases tried in the first quarter of 1999 and 2000) and the three interaction terms (which capture differential changes in the composition of the cases tried in the first quarter and those tried in the 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> quarters respectively). In particular, the coefficients associated with the interaction of the year and 2<sup>nd</sup> quarter dummies allow us to capture differential changes occurring during the transition period (starting the week when the reform was signed and ending the week when it entered into force). We also report the associated standard errors, which are clustered at the district court and individual levels.

While we do not find evidence that there was any change in the nature of the cases tried in the first and fourth quarters between 1999 and 2000 (columns 1 and 4), we find strong evidence suggesting that there was during the 2<sup>nd</sup> quarter of the year 2000 (column 2). Indeed, we find that the average annual earnings and number of days worked were significantly lower among defendants tried in the 2<sup>nd</sup> quarter of the year 2000. Similarly, defendants with a full-time job accounted for a smaller share of the individuals tried during this period. The magnitude of the differences is quite large: at the sample mean, they represent drops of 16 percentage points in earnings and 18 percentage points in the share of full-time employees during the 2<sup>nd</sup> semester of the year 2000. Overall, these suggest that wealthier individuals managed to postpone their case until after the reform more frequently than other defendants – presumably because they had access to better legal counsel. We also find that defendants who had at least one prior drunk-driving conviction in the previous two years accounted for a smaller share of the individuals tried in the 2<sup>nd</sup> quarter of the year 2000, while those who had at least one prior conviction for any non-drunk-driving crime accounted for a greater share. While these results are more complicated to interpret, they further suggest that the way drunk-driving cases were handled in district courts during the transition period generated differences in the nature of the defendants tried before and after the reform. Furthermore, because individuals with recent drunk-driving arrests were among the least likely to be considered eligible for benefitting from the new law,<sup>45</sup> it seems unlikely that the selection mechanism originated from an attempt by the district courts to restrict the set of cases tried before the reform to those whose outcome did not depend on the entering into force of the reform. Finally, some compositional changes are also observed for the third quarter (column 3), which

---

<sup>45</sup> Again, the reform did not apply systematically to offenders who had already been placed on probation for a drunk-driving crime more than once or to those who were on probation at the time of the crime for an alcohol-related crime.

merely reflect the fact that the number of cases remained lower than usual in the aftermath of the reform.

Generally, these findings question the degree of consistency with which drunk-driving cases were handled in district courts, as well as the level of fairness with which defendants were treated by the justice system during the transition period. From a methodological point of view, this also raises some questions with respect to the performance of some traditional quasi-experimental estimators in the context of this reform and other, similar ones – a point to which we will come back at the end of this paper.

## V. Identification strategy

In order to measure the causal impact of the reform on offenders' post-sentencing criminal activities and labor market attachment, we therefore opt for an alternative strategy which consists in comparing the individuals tried for a drunk-driving crime committed *before* the reform and differing only in their probability of receiving a custodial sentence. This allows us to bypass the selection problem which would arise if we were to compare individuals tried before and after the reform for reasons just detailed.<sup>48</sup>

Our approach relies on two features of the justice system which, combined together, create exogenous variation in the probability of receiving a custodial sentence for offenders arrested just before a reform introducing a more lenient legislation for the crime they committed. The first of these two features is the characteristic of Danish legislation mentioned earlier, which guarantees that defendants tried after a reform for a crime committed prior to it must be tried under the more lenient of the two laws (Danish criminal code, §3).<sup>49</sup> Hence, individuals tried after July 1<sup>st</sup>, 2000 for a drunk-driving crime committed before that date were tried under the new law. The second of these two features is the significant time gap between the moment a crime is committed and the moment the decision of justice is rendered by a district court (as further documented below). Together, these features ensure that the closer to the

---

<sup>48</sup> Traditional quasi-experimental estimators typically raise additional selection problems. In particular, one concern is that the entering into force of the new law might have been accompanied (at least for a time) by more frequent police controls to compensate for the reduction in the expected costs of the punishment by increasing the probability of being caught drunk driving. Moreover, another concern is that potential offenders might have modified their behavior around the time of the reform. For instance, they might have anticipated the above-mentioned increase in road traffic controls and behaved more carefully in the weeks following the entering into force of the reform, thus reducing the overall number of drunk-driving crimes. Furthermore, conditional on individuals internalizing changes in the legislation, the reform should also have induced a modification in the characteristics of the individuals arrested for a drunk-driving crime after the law was passed. Indeed, the lowering of the cost associated with drunk-driving crimes should mechanically have led a new range of individuals to commit drunk-driving crimes (those incurring lower benefits from committing a crime and/or higher costs if caught), thereby increasing the overall number of drunk-driving crimes.

<sup>49</sup> See Langsted et al. (2014) for an explanation of the functioning of that law (p.36).

reform a crime was committed, the more likely the offender was to be tried after the reform under the new law, and therefore to avoid prison.<sup>50</sup>

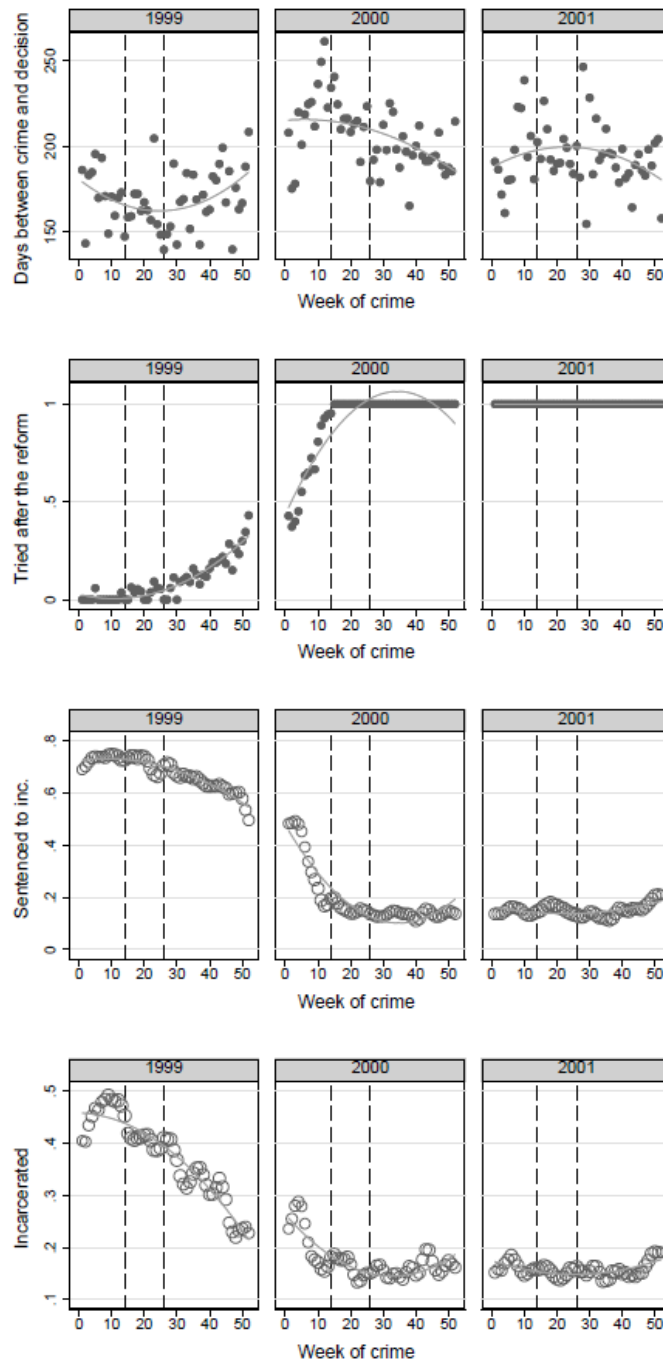
In *Figure 2*, we investigate the strength of this approach. In order to do so, the data is organized based on the week when the crime was committed (hereafter referred to as “*week of crime*”), instead of the week when the sentence was rendered, and the following indicators are depicted: a) the average time gap between the moment an alleged crime was committed and the moment the decision of justice was rendered by a district court by *week of crime*; b) the share of cases tried after July 1<sup>st</sup>, 2000 (the date when the reform officially entered into force) by *week of crime*; c) the share of defendants tried for an alleged drunk-driving crime who received a custodial sentence by *week of crime*; d) the share of defendants tried for an alleged drunk-driving crime who were actually incarcerated by *week of crime*.<sup>51</sup>

The evolution of the first two indicators provides a graphical motivation for this alternative approach. Around the time of the reform, the period of time between the moment when a prosecutor would press charges against an alleged drink-driver and the moment when a district court rendered its decision was substantial (and almost entirely driven by the case processing time in district courts). For drunk-driving crimes committed in 1999, this time gap was of 6 months on average. The closer to the reform the crime was committed, the more this figure increased. As a consequence, as individuals’ arrest date got closer to the reform within the 12-month period preceding it, an increasingly large share of them was tried after, under the new law. As for the last two indicators, their evolution suggests that, combined together, these two features introduced significant variation in the probability of receiving a custodial sentence among individuals tried for a drunk-driving crime committed in the 12-month period *preceding* the reform, based on the date of their crime. Indeed, the share of defendants who

---

<sup>50</sup> This identification strategy is loosely related to the one used in Drago et al. (2009), who used the Collective Clemency Bill passed by the Italian Parliament in July 2006 to measure the impact of suspended sentence length on recidivism. This reform reduced the length of the prison sentence of all inmates who had committed a crime before May 2, 2006. As a consequence, about 40 percent of the prison population of Italy were released from prison on August 1, 2006 under the condition that they would have to serve the remaining of their sentence if they were to commit another crime in the 5-year period following their release. Interestingly, the length of offenders’ suspended sentence varied depending on their prison entry date, which the researchers argued is exogenous and which they used to measure the impact of suspended sentence length on recidivism.

<sup>51</sup> For data confidentiality reasons, indicators c) and d) displayed in *Figure 2* are calculated as moving averages. For each year  $y$ , the value of these indicators (as displayed in *Figure 2*) are calculated as the average value of the indicators over years  $y-1$ ,  $y$ , and  $y+1$ . Furthermore, for any given week, the number of cases tried after the reform is normalized to 1 if the actual number of cases tried *after* is equal to or lower than 3 (in total, this normalization was carried out for 16 weeks), and the number of cases tried after the reform is normalized to 1 if the actual number of cases tried *before* is equal to or lower than 3 (in total, this normalization was carried out for 4 weeks).



**Figure 2 – Motivation for the instrumental variable approach**

This figure depicts the evolution of the following four indicators around the time of the reform: a) the average time gap between the moment a crime is committed and the moment the decision of justice is rendered by a district court by *week of crime*; b) the share of cases tried after July 1st, 2000 (the date when the reform officially entered into force) by *week of crime*; c) the share of defendants who received a custodial sentence by *week of crime*; d) the share of defendants who were actually incarcerated by *week of crime*. For each year, the first dotted vertical line marks the week when the law was signed (week 14) and the second one marks the week when it entered into force (week 26).

received a custodial sentence by week of crime started going down from July 1999 from slightly less than 80% to less than 20% right after the reform. The same pattern is observed for the share of defendants who were actually incarcerated following their trial – although the decrease starts earlier.

Because this approach compares individuals who committed their drunk-driving crime before the reform based on the date of their crime, the selection observed in the nature of defendants tried in the 3-month period preceding the entering into force of the reform must not affect the internal validity of the study results. However, by introducing variation in the probability of receiving a custodial sentence among offenders who should have been tried during this 3-month period (predominantly individuals arrested 6 to 9 months before the reform) based on their characteristics (e.g. their labor market attachment, prior convictions, etc.), this source of selection is likely to affect the set of individuals used by the instrumental variable approach to identify the effect of the reform, namely the compliers.

### *Sampling strategy*

Building on these observations, we compile a sample of defendants including all individuals charged for a drunk-driving crime committed in the 12-month period preceding the entering into force of the reform (between July 1<sup>st</sup>, 1999 and June 30<sup>th</sup>, 2000). We also include in our sample individuals charged for a drunk-driving crime committed 13 to 24 months before the entering into force of the reform (between July 1<sup>st</sup>, 1998 and June 30<sup>th</sup>, 1999), which allows us to control for seasonal variations using both a time trend, year and month fixed effects. Restricting our sample to defendants tried in one of the 84 district courts operating in the country at that time, we obtain a sample of 8,409 cases and defendants (including 8,012 distinct individuals<sup>52</sup>).

In *Table 3*, we provide a description of the characteristics of the defendants included in our sample. They are predominantly males in their late thirties. While close to 46.6% of them had a permanent full-time job at the end of the year preceding the date of their crime, 42.0% were not working. Strikingly, 18.3% of the defendants had already had at least one conviction for a drunk-driving crime in the previous 2 years, and 50.0% had already had at least one conviction for a drunk-driving crime before that. Moreover, few of them were in a partnership (28.2%), and defendants born abroad and

---

<sup>52</sup> In the few cases where an individual had allegedly committed more than one drunk-driving crime throughout the study period, keeping only the case associated with the first alleged drunk-driving crime yields results similar to those displayed below (results are available upon request).

descendants of immigrants represent 5.1% and 0.4% of the sample respectively – slightly less than their actual share in the overall population, which is 5.4% and 1.4% respectively.

### *Econometric specifications*

In order to report on the impact of the reform, we will first show the standard Ordinary-Least-Squares estimates (OLS) we obtain when estimating the following linear model:

$$y_{i,t} = \delta^{OLS} cust_i + X_i\beta + \mu_1 T_i + \mu_2 P_i + \mu_m + \mu_c + \varepsilon_i \quad (2)$$

where  $y_{i,t}$  is the outcome of interest for individual  $i$  measured at time  $t$ ;  $cust_i$  is a dummy variable indicating whether individual  $i$  received a custodial sentence as part of their trial;  $X_i$  is a vector including all variables in the conditioning set;  $T_i$  is a time trend indicating the number of days between the moment individual  $i$  committed their crime and the moment the reform entered into force on July, 1<sup>st</sup> 2000 (the unit for this variable is 100 days);  $P_i$  is a period dummy taking the value 1 if individual  $i$ 's crime was committed in the 12-month period preceding the reform and 0 if it was committed earlier;  $\mu_m$  and  $\mu_c$  are fixed effects indicating the month when individual  $i$  committed their crime and the district court where they were tried.

The coefficient  $\delta^{OLS}$  is the parameter of interest. However, for a number of reasons, the  $cust_i$  variable is likely to be endogenous in this specification. Indeed, as displayed in *Table 3*, the offenders who receive a custodial sentence and those who receive a non-custodial one differ significantly and, unless all differences across these two groups are controlled for (which seems unlikely to occur), OLS estimators are likely to yield biased estimates.

Next, we will report the reduced form estimates (RF) derived from the estimation of the following equation:

$$y_{i,t} = \delta^{RF} \overbrace{(P_i * T_i)}^{I_i} + X_i\beta + \mu_1 T_i + \mu_2 P_i + \mu_m + \mu_c + \varepsilon_i \quad (3)$$

where the  $(P_i * T_i)$  variable captures the differential effect of the  $T_i$  variable (indicating the number of days between the moment individual  $i$  committed their crime and the moment the reform entered into force) for crimes committed in the 12-month period preceding the day the reform entered into force, when compared to crimes committed in the 13 to 24 months before the reform. We use this variable as our instrument,  $I_i$ . *Figure 2* suggests that while there is no particular reason to expect  $\mu_1$  to be statistically different from 0 in this equation,  $\delta^{RF}$  should be if the nature of the sanctions imposed

on offenders before and after the reform has an impact on  $y_{i,t}$ , as the probability of receiving a custodial sentence is positively correlated with the time gap between the moment the crime was committed and the entering into force of the reform in the 12-month period preceding it.

Finally, we will report the instrumental variable (IV) estimates obtained by instrumenting  $cust_i$  by our instrument  $I_i$  using a Two-Stage-Least-Squares estimation procedure. Coefficient  $\delta^{IV}$  measures the impact of receiving a custodial sentence (as opposed to a non-custodial one) on the *compliers*, the subset of defendants for whom the time of their crime within the 12-month period preceding the entering into force of the reform had an impact on whether or not they received a custodial sentence – *i.e.* offenders who were sentenced to serve 1 to 60 days in prison.

#### *Variation in defendants' probability of being incarcerated*

In *Table 4*, we estimate the impact of having committed a drunk-driving crime closer to the date of the reform on the probability for a defendant to receive a custodial sentence (*Panel A*) and on the probability for a defendant to be actually incarcerated, which is captured by a proxy (*Panel B*). In order to do so, we regress the binary variable indicative of the trial outcome on our instrument and an increasingly exhaustive set of control variables. From column (1) to column (4), we enrich the set of control variables by adding the following covariates successively and incrementally: a time trend, period, month-of-crime and district court fixed effects (column 1), dummy variables indicative of the nature of the drunk-driving charge (column 2), information about the criminal case (column 3), and defendant characteristics (column 4).

As expected, we find that having committed a crime closer to the date of the reform substantially reduces the probability for a defendant to receive a custodial sentence when the crime was committed in the 12-month period preceding the entering into force of the reform (*Panel A*). Indeed, within that period, delaying their drunk-driving crime by 100 days would have reduced defendants' probability of receiving a custodial sentence by 19.9 percentage points. Furthermore, both the magnitude and significance level of these estimates are robust to the inclusion of covariates in the regression, suggesting that, in the 12-month period preceding the entering into force of the reform, the time gap between the day a defendant supposedly committed their crime and the moment the reform entered into force is independent of their characteristics and those of their case.

Similarly, we find that delaying their drunk-driving crime by 100 days would have reduced defendants' probability of being incarcerated by 5.5 percentage points in the 12-month period

preceding the entering into force of the reform (*Panel B*).<sup>53</sup> Both the magnitude and significance level of these estimates are also robust to the inclusion of covariates in the regression. The difference in the magnitude of the first-stage estimates displayed in *Panels A* and *B* can be explained by the implementation of a pardon scheme already mentioned above. As early as 1994, drink-drivers suffering from an alcohol abuse problem who received a prison sentence of no more than 60 days could apply to the Danish Prison and Probation Service so that their custodial sentence would be commuted to the sanction which was eventually generalized after the reform (a two-year probation period combined with mandatory participation in a yearlong rehabilitation program and a fine).<sup>54</sup> An important consequence is that the set of compliers on which this instrument variable relies to measure the relative impact of the custodial and non-custodial sentences is most likely less prone to severe alcoholic abuse problems than the average drink-driver.

Overall, these results suggest that the number of days separating the moment an alleged drunk-driving crime was committed in the 12-month period preceding the reform and the date of the reform may be used as an instrument to capture the impact of the reform (ITT estimates) and to measure the impact of substituting a non-custodial sentence for a custodial one (TOT estimates) – although on a selected subset of defendants.

#### *Instrument validity*

However, for this instrument to be valid, it has to meet the following standard conditions: independence, exclusion, and monotonicity. While the independence and exclusion assumptions are crucial for both ITT and TOT estimates, the monotonicity assumption only matters for TOT estimates.

The independence assumption implies that the instrument is independent of defendants' background characteristics and potential outcomes (once a time trend, period, month-of-crime and district court fixed effects are controlled for). In order to further investigate the validity of this assumption, we study whether or not defendants' pre-crime and case characteristics are correlated with the instrument. we do so by regressing each of the background variables displayed in the left column of *Table 3* on

---

<sup>53</sup> The impact of the time gap between the day the reform entered into force and the moment the crime was committed in the 12-month period preceding the reform is actually slightly underestimated when the proxy for whether or not a defendant was actually incarcerated following their trial is used as the dependent variable. Indeed, the share of offenders actually incarcerated started going down in the reference period, more than a year before the entering into force of the reform (as displayed in *Figure 2*). Regressing our proxy for whether or not a defendant was incarcerated on a constant, district court fixed effects, dummy variables indicative of the nature of the drunk-driving charge, information on the criminal case, and defendant characteristics yields a first-stage coefficient of 0.073 (and a standard error of 0.008).

<sup>54</sup> However, until the 2000 reform, drink-drivers had to apply to the Danish Prison and Probation Service to benefit from the pardon scheme.

the instrument, the time trend, period, month-of-crime and district court fixed effects. For each regression, we report the coefficients and standard errors associated with the instrument in *Table 3*. We find that the coefficients associated with the instrument are systematically small and largely insignificant, suggesting that the independence assumption is likely to be met.

The exclusion restriction implies that the timing of the crime *itself* does not have any direct impact on our outcome variables (defendants' crime and labor outcomes up to ten years after the completion of their trial). One concern is that the risk of recidivism and/or prospects of employment might vary across defendants based on the timing of their crime or the date of their sanction. However, the inclusion in our sample of individuals tried for a drunk-driving crime committed 13 to 24 months before the reform allows us to mitigate the consequences of this potential problem by controlling for trend and seasonality effects.

The monotonicity assumption implies that the probability of receiving a custodial sentence decreased for all offenders as their crime was committed closer to the reform in the 12 months preceding it. In order to investigate the validity of this assumption, we estimate the first-stage equation for various subgroups of the sample: males, females, individuals aged below 25, individuals aged above 25, individuals with prior drunk-driving convictions, individuals without any prior drunk-driving convictions, etc. We report the coefficients and standard errors associated with each of the subgroups in *Table A.3* (placed in the appendix). We find that the coefficients are all positive and statistically significant (as well as very similar in magnitude). This suggests that problems arising due to non-monotonicity are probably limited as well.

## **VI. Results**

We measure the impact of the reform on individuals' post-sentencing crime using the strategy described in the previous section. This allows us to understand the relative impact of the substituted custodial sentence and its non-custodial alternative on drink drivers' subsequent involvement in drunk-driving crimes and in any other types of crime. We also investigate their relative impact on offenders' labor market attachment to shed some light on a potentially important mechanism behind the effect observed on crime-related outcomes.

Our main estimates are displayed in *Table 5.A* (recidivism-related outcomes), *Table 5.B* (other crime outcomes), and in *Table 6* (labor market outcomes). In each table, OLS estimates are reported in

column (1), RF estimates in column (2), and IV estimates in columns (3) and (4). For IV estimates, columns (3) and (4) only differ in the set of covariates included in the estimated equation: in the former case, we only include the time trend, the period dummy (indicating whether or not a crime was committed in the 12-month period preceding the reform), month-of-crime and district court fixed effects; in the latter case, the whole conditioning set is added as well (again, this set provides information on the nature of the criminal case and defendants' background characteristics). Additional IV estimates are displayed in *Figures 3.a, 3.b, and 4* (the coefficients displayed are those which are obtained when all covariates are added to the estimated equation).

#### VI.A Relative impact on crime outcomes

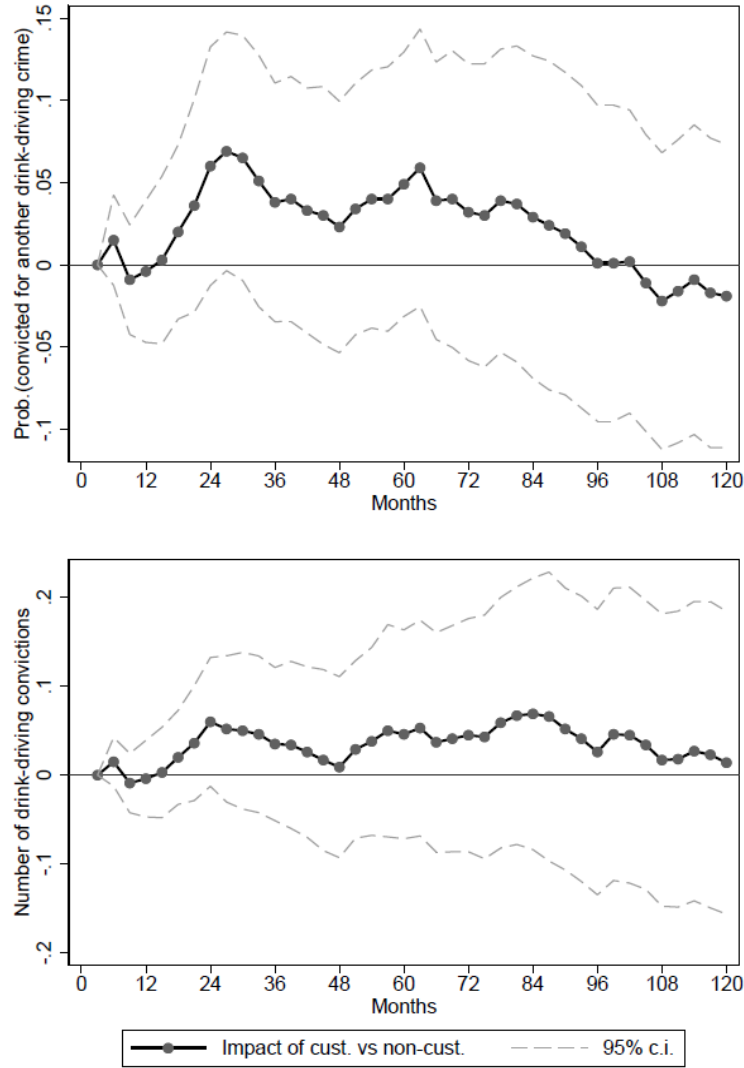
We assess the relative impact of a custodial sentence as compared to a non-custodial one on drink-drivers' post-sentencing criminal activities, as measured by their probability of being convicted of another drunk-driving crime and of any other crime (the *extensive* margin). For both types of crime, we also estimate the relative impact of the two sanctions on the number of crimes committed (the *intensive* margin). All of these outcomes are measured from the date the sentence was rendered.

In *Table 5.A*, we report on the relative impact of custodial and non-custodial sentences on drink-drivers' subsequent involvement in drunk-driving crimes. The goal is to understand whether or not drink drivers' recidivism remains a decreasing function of the severity of the sanction when the alternative to the custodial sanction is a non-custodial one – and not walking free without any sanction as in Hansen (2015). Interestingly, all estimates (OLS, RF, and IV) convey the same idea that custodial and non-custodial sentences are equally effective at preventing recidivism. Indeed, point estimates are close to 0 and all but two OLS coefficients fail to be statistically significant at the 10% level. This observation holds both at the intensive and extensive margins: the impact of the two sentences appears to be very similar on offenders' probability of committing any drunk-driving crimes and on the number of drunk-driving crimes they commit. Overall, these results suggest that the differential deterrent effect of the two sanctions was not important enough to spark any differences in drink-drivers' recidivism.

In *Tables 5.B*, we report on the relative impact of the two types of sanction on offenders' subsequent involvement in other types of crimes (non-drink driving crimes). The objective here is to understand whether or not drink-drivers can also be prone to the criminogenic effects custodial sentences have been found to have on other offenders. Interestingly, OLS estimates provide a picture consistent with

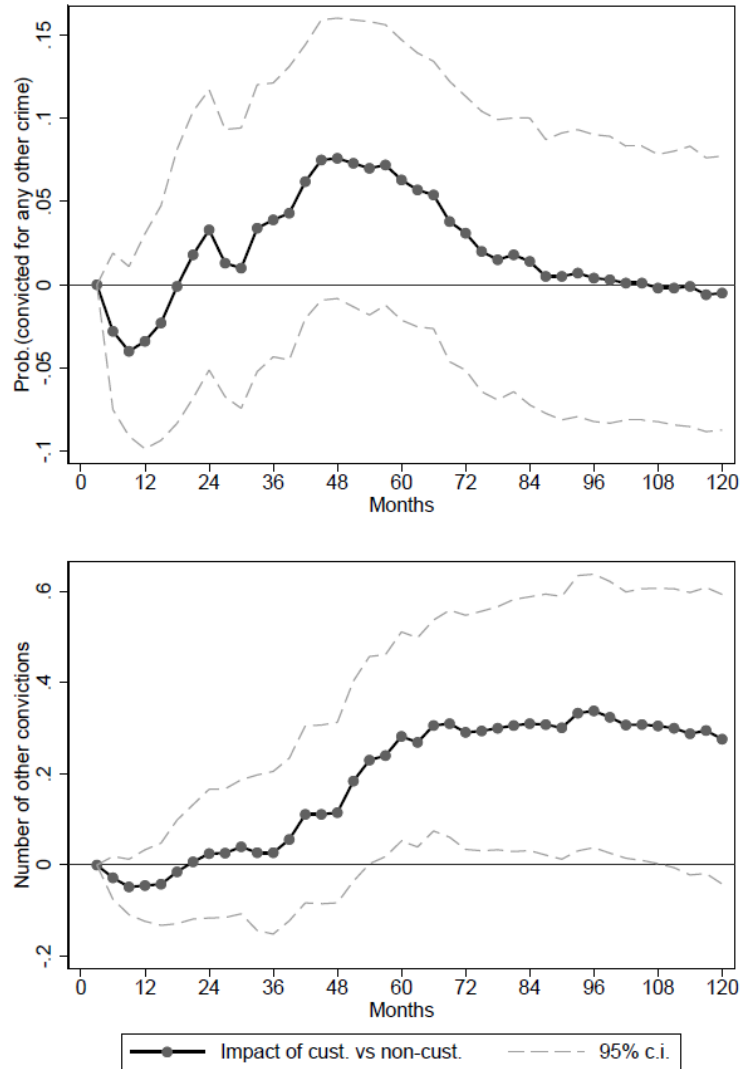
the deterrence theory, as they suggest that there is a negative relationship between receiving a custodial sentence and these other crime outcomes, both at the extensive and intensive margins. Indeed, most of these coefficients are negative and statistically significant at the 5% level, suggesting that custodial sentences decrease both the probability for a drink-driver to be subsequently convicted of any other crime, as well as the number of such crimes they commit. For instance, they suggest that custodial sentences decrease the probability for an offender to commit any other crimes within 5 years by 4.3 percentage points (representing a 9.7% decrease at the mean) and reduces the number of such crimes they commit by 0.127 crime (representing a 13.6% decrease at the mean) – both results are statistically significant at the 1% level. However, as already mentioned earlier, these estimates are likely to be biased given that it is unlikely that our specification controls for all differences between the offenders who received a custodial sentence and those who received a non-custodial one.

As a matter of fact, RF and IV estimates provide a completely different picture, which suggests that while custodial and non-custodial sentences appear to be equally effective at preventing individuals from committing at least one other crime (the extensive margin), custodial sentences increase the average number of crimes offenders commit. For instance, our results indicate that custodial sentences increase the number of non-drunk-driving crimes they commit within 5 years by 0.282 crime (representing a 30.1% increase at the mean) – this result is statistically significant at the 5% level. However, the difference between the average number of other crimes committed by offenders who received a custodial sentence and by those who received a non-custodial one remains stable throughout the rest of the study period (5 to 10 years after the decision of justice). This suggests that the relative criminogenic effect of custodial sentences is only transitory, materializing in the years following the incarceration.



**Figure 3.a – Impact of custodial vs. non-custodial sentences on drunk-driving crimes**

This figure depicts the cumulative impact of a custodial sentence (as measured by our IV estimates) on the following outcomes: a) the probability of being convicted of a drunk-driving crime; and b) the number of drunk-driving crimes committed. We measure crime outcomes every 3 months from the date when the drunk-driving case was settled in court.

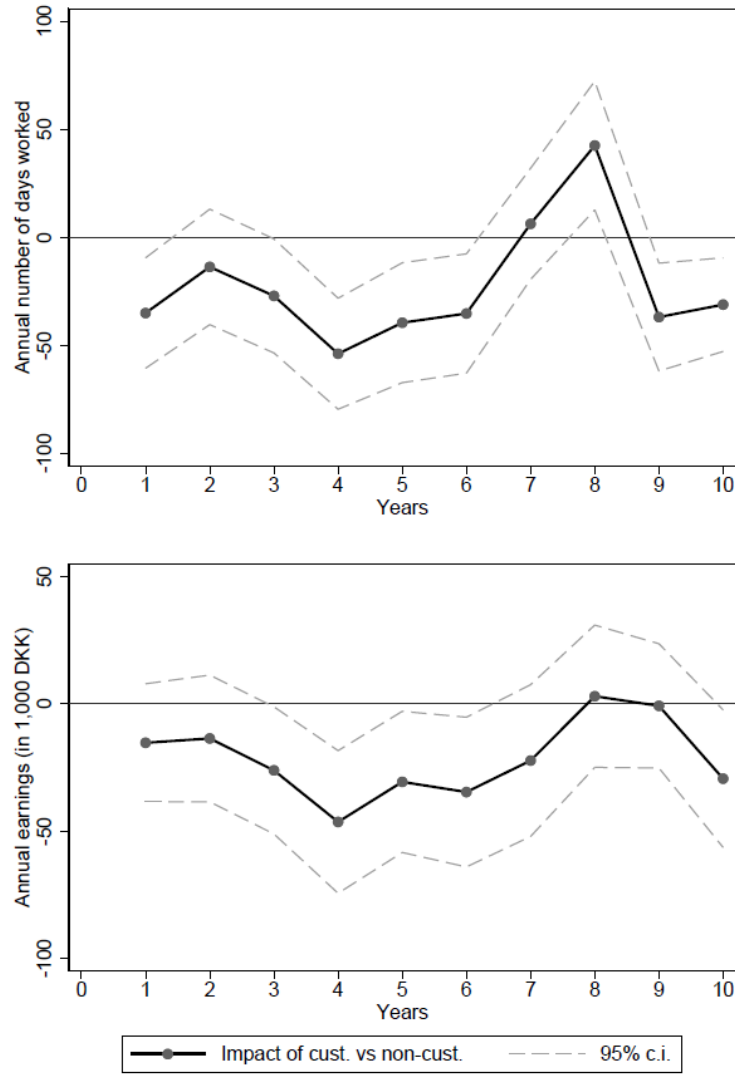


**Figure 3.b – Impact of custodial vs. non-custodial sentences on other crimes**  
 This figure depicts the cumulative impact of custodial sentence (as measured by our IV estimates) on the following outcomes: a) the probability of being convicted of any other crime; and b) the number of other crimes committed. We measure crime outcomes every 3 months from the date when the drunk-driving case was settled in court.

## VI.B Relative impact on labor market outcomes

In *Table 6*, we report on the relative impact of custodial and non-custodial sentences on two outcomes documenting offenders' labor market attachment: their annual number of days worked (*Panel A*), and their annual earnings before tax and any social contributions (*Panel B*). The aim here is first to understand the relative impact of these two sanctions on a key aspect of offenders' lives, as well as to shed some light on the mechanisms behind the effects found on crime-related outcomes. First of all, it is interesting to note that both the number of days worked and earnings decrease from the moment the decision of justice is rendered. For instance, the number of days worked decreases from 129.5 days at the end of the year following the trial to 69.2 days 10 years after. While part of this evolution is mechanically driven by a subset of individuals entering retirement, it also suggests that being convicted of a drunk-driving crime adversely affects offenders' labor market attachment (irrespective of whether or not the sentence is a custodial or a non-custodial one).

While OLS estimates suggest that custodial sentences do not have any impact on any of the two labor market outcomes, RF and IV estimates suggest that custodial sentences significantly weaken individuals' labor market attachment and do so on a long-term basis (compared to non-custodial ones). Indeed, despite the limited length of the incarceration spells, offenders who receive a custodial sentence work significantly fewer days every year after they receive their sentence than those who only receive non-custodial sentences. Moreover, this result holds for almost every year following the decision of justice. As a consequence, they also earn much less every year. For instance, 5 years after the decision of justice, individuals who received a custodial sentence work on average 40.6 days less than individuals who received a non-custodial one – representing a 34.5% decrease at the mean. They also earn 31,519 kroners less than individuals who received a non-custodial sentence – representing a 24.4% decrease at the mean.



**Figure 4 – Impact of custodial vs. non-custodial sentences on labor outcomes**

This figure depicts the impact of custodial sentences (as measured by our IV estimates) on the following two outcomes: a) annual number of days worked; b) annual earnings before tax and any social contributions. We measure labor outcomes every year starting at the end of the year following the date when the drunk-driving case was settled in court.

Also, it is interesting to observe that differences in individuals' labor market attachment between offenders who received a custodial sentence and those who received a non-custodial one are negative but remain small in the first two years following the decision of justice. However, they become more important from the 3<sup>rd</sup> year, which marks the end of the probation period for most of the offenders who received a non-custodial sentence and were placed on probation. This suggests that non-custodial sentences can constitute an important disruption in individuals' lives, which negatively affects their integration on the labor market, but that its longer-term effects are relatively less serious than those associated with a custodial sentence (even a short one).

Finally, incarcerated offenders' labor market attachment seems to improve relatively towards the end of the period (around 8 years after the decision of justice) before worsening again at the end of the period studied. While this pattern remains puzzling, it is interesting to note that, for most offenders in the sample, the 9<sup>th</sup> and 10<sup>th</sup> years after the decision of justice correspond to 2009 and 2010, which were marked by a financial crisis and a sharp rise in the unemployment rate.<sup>55</sup> In turn, although further research is needed before any definitive conclusions can be drawn, these observations seem to suggest that while the relative negative impact of custodial sentences on offenders' employment prospects might fade out with time, they remain more liable to economic downturns.

## **VIII. Conclusion**

In this article, we study a large-scale reform of the Danish legislation implemented in 2000, whereby a custodial sentence (jail time) was replaced by a non-custodial one (a probation period, combined with a fine and either mandatory participation in a yearlong rehabilitation program or community work) for most drunk-driving crimes. The study reaches several conclusions.

First, we find evidence suggesting that salient contextual changes (such as a legislative reform, a program scale-up, etc.) can be anticipated by their stakeholders, who in turn can modify their behavior in line with their best interest. In the case of the change in the drunk-driving legislation studied here, we show that the way drunk-driving cases were handled in district courts changed drastically in the months preceding the entering into force of the reform (from the moment the law was signed) and

---

<sup>55</sup> The unemployment rate evolved as follows in Denmark between 2007 and 2011: 3.8% in 2007, 3.4% in 2008, 6.0% in 2009, 7.5% in 2010, and 7.6% in 2011.

that a large share of the cases which should have been tried before the reform was actually tried after. This implies that a group of offenders who should have been tried before the reform were tried after and, as a consequence, avoided prison. Furthermore, we show that the identity of the individuals who had their case postponed was not random and that, for instance, wealthier defendants were more likely to have their trial delayed until after the reform. From a policy perspective, it suggests that it would be advisable to synchronize the passing and entering into force of a new law or, whenever possible, to monitor more closely the way cases are handled in such periods so as not to introduce any avoidable source of inequities in the justice system.

Second, we show that the substitution of non-custodial sentences for custodial ones, which occurred as part of the reform studied in this article, did not have any impact on offenders' probability of committing another drunk-driving crime or on the number of drunk-driving crimes committed up to 10 years after the decision of justice. However, it managed to reduce the average number of other types of crimes committed while holding the stock of reoffenders constant. Furthermore, we find that, compared to custodial sentences, non-custodial sentences significantly improve offenders' labor market attachment as they increase both their annual number of days worked and earnings. Although the peculiar nature of the compliers on which the impact of the reform is estimated rules out any definitive conclusions with respect to the relative effect of custodial and non-custodial sentences, our findings suggest that more severe punishments do not always reduce the probability for drink-drivers to recidivate and that non-custodial sentences can be at least as effective as custodial ones (and much cheaper – see Nielsen and Kyvsgaard (2007), and Henrichson and Delaney (2012)) in preventing re-offense.

## References

- Abrams, D. S., Bertrand, M., & Mullainathan, S. (2012). Do judges vary in their treatment of race?. *The Journal of Legal Studies*, 41(2), 347-383.
- Aizer, A., & Doyle, J. J. (2015). "Juvenile Incarceration, Human Capital and Future Crime: Evidence from Randomly-Assigned Judges." *The Quarterly Journal of Economics*, qjv003.
- Anwar, S., Bayer, P., & Hjalmarsson, R. (2014). The role of age in jury selection and trial outcomes. *The Journal of Law and Economics*, 57(4), 1001-1030.
- Anwar, S., Bayer, P., & Hjalmarsson, R. (2012). The Impact of Jury Race in Criminal Trials. *The Quarterly Journal of Economics*, qjs014.
- Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, 47-57.
- Becker, G. S. (1968). Crime and punishment: An economic approach. *Essays in the Economics of Crime and Punishment* (pp. 1-54). UMI.
- Blundell, R., & Costa Dias, M. (2000). Evaluation methods for non-experimental data. *Fiscal studies*, 21(4), 427-468.
- Calinescu, T., & Adninaite, D. (2018). Progress in reducing drink driving in Europe.
- Chalfin, A., & McCrary, J. (2017). Criminal deterrence: A review of the literature. *Journal of Economic Literature*, 55(1), 5-48.
- Chou, S. P., Dawson, D. A., Stinson, F. S., Huang, B., Pickering, R. P., Zhou, Y., & Grant, B. F. (2006). The prevalence of drinking and driving in the United States, 2001–2002: results from the national epidemiological survey on alcohol and related conditions. *Drug and alcohol dependence*, 83(2), 137-146.
- Clausen, S. (2007). Samfundstjeneste virker det? *Djøf/Jurist-og Økonomforbundet*.
- Cohen, A., & Yang, C. S. (2019). Judicial Politics and Sentencing Decisions. *American Economic Journal: Economic Policy*, 11 (1): 160-91.

- Cullen, F. T., Jonson, C. L., & Nagin, D. S. (2011). Prisons do not reduce recidivism: The high cost of ignoring science. *The Prison Journal*, 91(3\_suppl), 48S-65S.
- Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17), 6889-6892.
- Drago, F., Galbiati, R., & Vertova, P. (2009). The deterrent effects of prison: Evidence from a natural experiment. *Journal of political Economy*, 117(2), 257-280.
- European Commission (2015), Alcohol, Directorate General for Transport, <https://goo.gl/q1jCS8>
- Hansen, B. (2015). Punishment and Deterrence: Evidence from Drunk Driving. *The American Economic Review*, 105(4), 1581-1617.
- Heckman, J. J., LaLonde, R. J., & Smith, J. A. (1999). The economics and econometrics of active labor market programs. In *Handbook of labor economics* (Vol. 3, pp. 1865-2097). Elsevier.
- Heckman, J. J., & Smith, J. A. (1999). The Pre-programme Earnings Dip and the Determinants of Participation in a Social Programme. Implications for Simple Programme Evaluation Strategies. *The Economic Journal*, 109(457), 313-348.
- Henrichson, C. & Delaney, R. (2012). *The Price of Prisons: What Incarceration Costs Taxpayers*. New York: Vera Institute of Justice.
- Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of economic literature*, 47(1), 5-86.
- Kramp, P. et al. (1990). Alkoholvaner blandt kriminalforsorgens klientel. Kriminalpolitisk Forskningsgruppe, forskningsrapport nr. 31, Justitsministeriet, København, 1990.
- Kriminal Forsorgen Årsrapporten (1998). Årsrapport 1998.
- Kriminal Forsorgen Årsrapporten (1999). Årsrapport 1999.
- Kyvsgaard, B. (2004). Youth justice in Denmark. *Crime and justice*, 349-390.
- Langsted, L. B., Garde, P., & Greve, V. (2014). *Criminal law in Denmark*. Kluwer Law International.

- Lapham, S. C., Smith, E., C'de Baca, J., Chang, I., Skipper, B. J., Baum, G., & Hunt, W. C. (2001). Prevalence of psychiatric disorders among persons convicted of driving while impaired. *Archives of General Psychiatry*, 58(10), 943-949.
- Lappi-Seppälä, T. (2007). Penal policy in Scandinavia. *Crime and justice*, 36(1), 217-295.
- Nagin, D. S., & Pogarsky, G. (2001). Integrating celerity, impulsivity, and extralegal sanction threats into a model of general deterrence: Theory and evidence. *Criminology*, 39(4), 865-892.
- National Center for Statistics and Analysis. (2017). Alcohol-impaired driving: 2016 data (Traffic Safety Facts. Report No.HS 812 450). Washington, DC: National Highway Traffic Safety Administration.
- Nielsen, R. C., & Kyvsgaard, B. (2007). Alkoholistbehandling. En effektevaluering.
- Philippe, A., & Ouss, A. (2018). “No hatred or malice, fear or affection”: Media and sentencing. *Journal of Political Economy*, 126(5).
- Pratt, J. (2008). Scandinavian exceptionalism in an era of penal excess part I: the nature and roots of scandinavian exceptionalism. *British Journal of Criminology*, 48(2), 119-137.
- Pratt, J., & Eriksson, A. (2011). ‘Mr. Larsson is walking out again’. The origins and development of Scandinavian prison systems. *Australian & New Zealand Journal of Criminology*, 44(1), 7-23.
- Vidmar, N. (2011). The psychology of trial judging. *Current Directions in Psychological Science*, 20(1), 58-62.
- Vingilis, E. (1983). Drinking Drivers and Alcoholics Are They From the Same Population?. In *Research advances in alcohol and drug problems* (pp. 299-342). Springer US.
- Vissers, L. H. (2017). Alcohol-related road casualties in official crash statistics. Paris, International Transport Forum ITF, 2017, 55 p., ref.; International Traffic Safety Data and Analysis Group IRTAD Research Report.
- Ward, K., Longaker, A. J., Williams, J., Naylor, A., Rose, C. A., & Simpson, C. G. (2013). Incarceration within American and Nordic prisons: Comparison of national and international policies. *The International Journal of Research and Practice on Student Engagement*, 1(1), 36-47.

Western, B., Kling, J. R., & Weiman, D. F. (2001). The labor market consequences of incarceration. *Crime & delinquency*, 47(3), 410-427.

World Health Organization. (2007). Drinking and driving: a road safety manual for decision-makers and practitioners. *Drinking and driving: a road safety manual for decision-makers and practitioners*.

World Health Organization. (2018). Global status report on road safety 2018.

Yu, J., Evans, P. C., & Clark, L. P. (2006). Alcohol addiction and perceived sanction risks: Deterring drinking drivers. *Journal of Criminal Justice*, 34(2), 165-174.

## Tables

**Table 1: Drunk-driving trial outcomes in cases tried in 1999**

	<b>Drink-driving crimes tried in 1999</b>		
	<b>#Obs.</b>	<b>Mean</b>	<b>S.d.</b>
No sanction	4,249	0.013	0.115
Prison	4,249	0.726	0.446
Prison, cond.	4,249	0.007	0.085
Prison, cond. (length in days)	4,249	0.199	3.169
Prison, uncond.	4,249	0.719	0.449
Prison, uncond. (length in days)	4,249	18.129	25.402
<i>Between 1 and 60 days</i>	4,249	0.685	0.465
<i>Over 60 days</i>	4,249	0.052	0.221
Imprisoned	4,249	0.437	0.496
Community work	4,249	0.000	0.000
Fine	4,249	0.278	0.448
Fine amount	4,249	1624.441	3109.778
Driv. lic. suspended	4,249	0.308	0.462
Driv. lic. suspended (length in days)	4,249	1.851	2.774
Appeal	4,249	0.017	0.128

*Notes:* In this table, I provide a description of the trial outcomes of the drink-driving cases tried in 1999.

**Table 2: Selection in the nature of the cases tried before, during, and after the transition period**

Variables	Whole sample			Year 2000		Weeks 14-26		Weeks 27-39		Weeks 40-52	
	#Obs.	Mean	S.d.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
<i>Criminal background</i>											
Nber of past convict.	7,894	4.514	4.772	0.092	0.828	0.659	0.349 *	-0.345	0.299	0.137	0.281
<i>0-2 years priors</i>											
Any past convict.	7,894	0.407	0.491	0.018	0.073	0.027	0.033	-0.010	0.028	-0.042	0.031
Nber of past convict.	7,894	0.647	0.962	0.019	0.184	0.064	0.067	0.001	0.062	-0.081	0.062
Any past DD convict.	7,894	0.177	0.382	0.003	0.056	-0.041	0.020 **	-0.021	0.024	-0.007	0.019
Nber of past DD convict.	7,894	0.196	0.441	0.012	0.063	-0.047	0.023 **	-0.019	0.028	-0.016	0.019
<i>&gt;2 years priors</i>											
Any past convict.	7,894	0.793	0.405	-0.004	0.082	0.060	0.027 **	-0.007	0.029	0.044	0.020 **
Nber of past convict.	7,894	3.869	4.378	0.076	0.724	0.594	0.306 *	-0.348	0.273	0.214	0.248
Any past DD convict.	7,894	0.491	0.500	-0.042	0.078	0.000	0.031	-0.070	0.043	0.032	0.031
Nber of past DD convict.	7,894	0.849	1.153	0.014	0.177	0.007	0.074	-0.201	0.091 **	0.050	0.067
<i>Attachment to the labor market</i>											
Active	7,842	0.756	0.429	0.019	0.067	-0.019	0.027	-0.001	0.030	-0.027	0.024
Employed	7,842	0.645	0.479	0.001	0.083	-0.024	0.031	-0.012	0.036	-0.028	0.028
Unemployed	7,842	0.112	0.315	0.018	0.050	0.005	0.018	0.011	0.020	0.001	0.018
Full-time employ. (>29h/week)	7,894	0.469	0.499	0.072	0.076	-0.086	0.032 ***	-0.086	0.036 **	-0.031	0.026
Part time employ. (15-29h/week)	7,894	0.053	0.223	0.026	0.037	0.048	0.018 ***	0.043	0.016 ***	0.009	0.015
Limited empl. <15h/week)	7,894	0.029	0.169	-0.039	0.025	-0.004	0.010	0.007	0.011	0.001	0.009
Employ. w/ partial unempl. during the year	7,894	0.009	0.096	-0.015	0.017	0.006	0.006	0.004	0.007	0.005	0.006
Unknown	7,894	0.022	0.146	0.005	0.024	0.013	0.010	0.012	0.008	0.000	0.009
None	7,894	0.418	0.493	-0.049	0.082	0.023	0.030	0.020	0.038	0.016	0.026
Nber of days worked	7,894	149.648	162.488	19.187	26.033	-21.613	9.736 **	-8.333	12.244	-12.065	9.599
Earnings (in 1,000 DKK)	7,894	160.407	159.095	13.006	24.795	-25.605	9.994 **	-18.280	13.221	-10.984	11.020

*Notes:* In this table, I describe the characteristics (mean and standard deviation) of the defendants tried between January 1st, 1999 and December 31st 2000. I also investigate whether or not the characteristics of the offenders tried in each quarter of the year 2000 remained stable. In order to do so, I regress each of the variables displayed in the left column of this table on a constant, a time trend, a dummy variable indicating when a defendant was tried in 2000 (as opposed to 1999), three dummy variables indicating when a defendant was tried in the 2nd quarter (weeks 14 to 26), the third (weeks 27 to 39), or the fourth (weeks 40 to 52) of either 1999 or 2000, as well as the interaction of the year and quarter dummies. I report the coefficient and standard error associated with the year dummy and the interaction variables. Standard errors are clustered at the district court and individual levels. Significance levels: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table 3: Sample description and balance checks**

Variables	Whole sample			Instrument	
	#Obs.	Mean	S.d.	Coeff.	s.e.
<i>Defendant characteristics</i>					
Female	8,409	0.084	0.278	0.001	0.006
Juvenile, when the crime was committed	8,409	0.003	0.058	0.001	0.001
Age when the decision was rendered	8,409	38.291	11.068	-0.260	0.227
<i>Immigration status</i>					
Immigrant	8,409	0.051	0.220	-0.007	0.004
Descendant	8,409	0.004	0.063	0.000	0.001
Rest of the population	8,409	0.945	0.228	0.007	0.004
<i>Family status</i>					
Single	8,409	0.498	0.500	-0.001	0.011
In partnership	8,409	0.282	0.450	0.003	0.009
Separated	8,409	0.194	0.396	0.002	0.007
Widow	8,409	0.014	0.117	0.001	0.003
Unknown	8,409	0.012	0.109	-0.006	0.003 **
<i>Education status</i>					
Months of educ. required to obtain diploma	8,100	141.971	30.709	0.109	0.675
Primary education	8,409	0.527	0.499	-0.004	0.011
Secondary education	8,409	0.379	0.485	0.001	0.010
Higher education	8,409	0.059	0.235	0.001	0.005
Unknown	8,409	0.035	0.185	0.003	0.004
<i>Attachment to the labor market</i>					
Earnings (in 1,000 DKK)	8,409	160.297	158.905	2.482	3.500
Nber of days worked	8,409	149.815	162.686	1.518	3.535
Full-time employ. (>29h/week)	8,409	0.466	0.499	0.013	0.010
Part time employ. (15-29h/week)	8,409	0.053	0.224	0.001	0.005
Limited empl. <15h/week)	8,409	0.030	0.171	-0.003	0.003
Employ. w/ partial unempl. during the year	8,409	0.010	0.100	-0.003	0.002
Unknown	8,409	0.022	0.146	0.004	0.003
None	8,409	0.420	0.494	-0.013	0.010
<i>Criminal case characteristics</i>					
Mixed court trial	8,409	0.033	0.179	-0.005	0.003
Nber of past convict.	8,409	4.587	4.847	-0.020	0.094
<i>0-2 years priors</i>					
Any past convict.	8,409	0.414	0.493	0.004	0.011
Nber of past convict.	8,409	0.659	0.966	0.008	0.020
Any past DD convict.	8,409	0.183	0.387	0.002	0.008
Nber of past DD convict.	8,409	0.202	0.446	0.002	0.009
<i>&gt;2 years priors</i>					
Any past convict.	8,409	0.794	0.405	0.004	0.008
Nber of past convict.	8,409	3.930	4.453	-0.027	0.083
Any past DD convict.	8,409	0.500	0.500	0.006	0.010
Nber of past DD convict.	8,409	0.861	1.155	0.020	0.023

*Notes:* In this table, I describe the characteristics (mean and standard deviation) of the set of defendants included in my sample and report how defendants' characteristics are correlated with the instrument. The estimates describing the differential characteristics are obtained by regressing each of the variables displayed in the left column of the table on a constant, my instrument, a trend, a dummy variable indicative of whether or not the crime was committed in the 12 month period preceding the entering into force of the reform, and month of crime and district court fixed effects. Standard errors are clustered at the district court and individual levels. Significance levels: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table 4: Reform and custodial sentence**

	(1)	(2)	(3)	(4)
<i>Section A: Probability of receiving a custodial sentence</i>				
Instrument	0.200*** (0.011)	0.199*** (0.010)	0.199*** (0.010)	0.199*** (0.010)
Mean	0.574	0.574	0.574	0.574
Observations	8,409	8,409	8,409	8,409
R-squared	0.217	0.234	0.281	0.293
<i>Section B: Probability of being incarcerated</i>				
Instrument	0.055*** (0.009)	0.055*** (0.008)	0.055*** (0.008)	0.056*** (0.008)
Mean	0.346	0.346	0.346	0.346
Observations	8,409	8,409	8,409	8,409
R-squared	0.077	0.085	0.094	0.106
Trend	YES	YES	YES	YES
Period FE	YES	YES	YES	YES
Month of crime FE	YES	YES	YES	YES
District court FE	YES	YES	YES	YES
Charge FE	NO	YES	YES	YES
Case charact.	NO	NO	YES	YES
Def. charact.	NO	NO	NO	YES

*Notes:* In this table, I measure impact of my instrument (the number of days between the moment an individual committed their crime and the time the reform entered into force, in 100 days) on the probability for them to receive a custodial sentence (*Panel A*) and to be actually incarcerated (*Panel B*). For each of these outcomes, I regress the dependent variable on my instrument and an increasingly exhaustive set of covariates: in column (1), a time trend, a dummy variable indicative of whether or not the crime was committed in the 12 month period preceding the entering into force of the reform, and month of crime and district court fixed effects are added to the regression; in column (2) dummy variables indicative of the nature of the charge are added as well; in column (3), I add information on the criminal case; in column (4), I add defendant characteristics. Standard errors are clustered at the district court and individual levels. Significance levels: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table 5.A: Impact on drunk-driving crimes**

	Whole sample			OLS (1)		RF (2)		IV (3) (4)			
	N	Mean	S.d.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
<i>Panel A: Probability of being convicted again for drink-driving</i>											
Within 6 months	8,409	0.014	0.119	0.007	0.003 **	0.003	0.003	0.015	0.013	0.015	0.014
Within 1 year	8,409	0.045	0.207	0.005	0.006	-0.001	0.004	-0.007	0.021	-0.004	0.022
Within 2 years	8,409	0.105	0.307	0.009	0.009	0.012	0.008	0.055	0.037	0.060	0.037
Within 5 years	8,409	0.242	0.428	0.000	0.011	0.010	0.009	0.050	0.041	0.049	0.041
Within 10 years	8,409	0.355	0.478	0.010	0.012	-0.004	0.010	-0.009	0.048	-0.020	0.047
<i>Panel B: Number of drink-driving convictions</i>											
Within 6 months	8,409	0.014	0.119	0.007	0.003 **	0.003	0.003	0.015	0.013	0.015	0.014
Within 1 year	8,409	0.045	0.207	0.005	0.006	-0.001	0.004	-0.007	0.021	-0.004	0.022
Within 2 years	8,409	0.105	0.307	0.009	0.009	0.012	0.008	0.055	0.037	0.060	0.037
Within 5 years	8,409	0.316	0.628	0.015	0.018	0.009	0.013	0.042	0.060	0.046	0.060
Within 10 years	8,409	0.539	0.882	0.047	0.024 **	0.003	0.018	0.022	0.091	0.013	0.087
Trend				YES		YES		YES		YES	
Period FE				YES		YES		YES		YES	
Month-of-crime FE				YES		YES		YES		YES	
District court FE				YES		YES		YES		YES	
Additional Cov.				YES		YES		NO		YES	
F-test	368.518										

*Notes:* In this table, I report Ordinary Least Squares (OLS) estimates derived from the regression of my outcome variable on a dummy variable indicating whether or not individual  $i$  received a custodial sentence (equation 2); Reduced-Form (RF) estimates derived from the regression of my outcome variable on my instrument (equation 3); finally, I report the Instrumental Variables (IV) estimates obtained by instrumenting the dummy variable indicating whether or not individual  $i$  received a custodial sentence by my instrument. For each category of estimates, I report estimates obtained when a trend, a dummy variable indicative of whether or not the crime was committed in the 12 month-period preceding the entering into force of the reform, month of crime and district court fixed effects, and the whole conditioning set are added to the estimated equation (columns 1, 2, and 4). For IV estimates, I also report the results obtained when the conditioning set is not added to the estimated equation (column 3). Standard errors are clustered at the district court and individual levels. Significance levels: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

**Table 5.B: Impact on other crimes**

	Whole sample			OLS (1)		RF (2)		IV (3)		IV (4)	
	N	Mean	S.d.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
<i>Panel A: Probability of being convicted again for any crime other than drink-driving</i>											
Within 6 months	8,409	0.050	0.218	-0.008	0.005 *	-0.006	0.005	-0.026	0.024	-0.028	0.024
Within 1 year	8,409	0.133	0.339	-0.019	0.009 **	-0.007	0.007	-0.032	0.036	-0.034	0.033
Within 2 years	8,409	0.264	0.441	-0.032	0.011 ***	0.007	0.009	0.038	0.045	0.033	0.043
Within 5 years	8,409	0.444	0.497	-0.043	0.011 ***	0.013	0.009	0.073	0.049	0.063	0.043
Within 10 years	8,409	0.561	0.496	-0.043	0.010 ***	-0.001	0.009	0.008	0.048	-0.007	0.042
<i>Panel B: Number of convictions other than drink-driving convictions</i>											
Within 6 months	8,409	0.050	0.218	-0.008	0.005 *	-0.006	0.005	-0.026	0.024	-0.028	0.024
Within 1 year	8,409	0.150	0.403	-0.025	0.010 **	-0.009	0.008	-0.043	0.045	-0.045	0.040
Within 2 years	8,409	0.375	0.733	-0.052	0.020 ***	0.005	0.015	0.032	0.078	0.025	0.072
Within 5 years	8,409	0.936	1.477	-0.127	0.039 ***	0.056	0.024 **	0.291	0.146 **	0.282	0.117 **
Within 10 years	8,409	1.612	2.326	-0.238	0.051 ***	0.055	0.033 *	0.298	0.217	0.276	0.162 *
Trend				YES		YES		YES		YES	
Period FE				YES		YES		YES		YES	
Month-of-crime FE				YES		YES		YES		YES	
District court FE				YES		YES		YES		YES	
Additional Cov.				YES		YES		NO		YES	
F-test											368.518

*Notes:* In this table, I report Ordinary Least Squares (OLS) estimates derived from the regression of my outcome variable on a dummy variable indicating whether or not individual  $i$  received a custodial sentence (equation 2); Reduced-Form (RF) estimates derived from the regression of my outcome variable on my instrument (equation 3); finally, I report the Instrumental Variables (IV) estimates obtained by instrumenting the dummy variable indicating whether or not individual  $i$  received a custodial sentence by my instrument. For each category of estimates, I report estimates obtained when a trend, a dummy variable indicative of whether or not the crime was committed in the 12 month-period preceding the entering into force of the reform, month of crime and district court fixed effects, and the whole conditioning set are added to the estimated equation (columns 1, 2, and 4). For IV estimates, I also report the results obtained when the conditioning set is not added to the estimated equation (column 3). Standard errors are clustered at the district court and individual levels. Significance levels: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

**Table 6: Impact on annual number of days worked and earnings**

	Whole sample			OLS (1)		RF (2)		IV (3) (4)			
	N	Mean	S.d.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
<i>Panel A: Annual number of days worked</i>											
Within 1 year	8,409	129.501	159.780	-5.216	3.419	-6.978	2.620 ***	-23.492	14.992	-34.987	13.023 ***
Within 2 years	8,409	128.117	160.621	0.215	3.549	-2.721	2.778	-2.579	14.170	-13.642	13.620
Within 3 years	8,409	123.293	160.320	-1.231	4.275	-5.398	2.735 **	-16.357	14.712	-27.063	13.456 **
Within 4 years	8,409	118.026	159.195	-0.650	3.415	-10.730	2.722 ***	-42.153	14.909 ***	-53.798	13.095 ***
Within 5 years	8,409	117.703	159.507	-2.854	3.823	-7.868	2.910 ***	-27.601	15.336 *	-39.449	14.141 ***
Within 6 years	8,409	120.342	159.473	-4.455	4.148	-7.020	2.975 **	-23.128	15.808	-35.195	14.075 **
Within 7 years	8,409	120.407	159.948	-1.455	3.978	1.253	2.646	18.616	14.915	6.281	13.112
Within 8 years	8,409	100.011	153.577	1.535	3.648	8.474	3.156 ***	52.246	16.814 ***	42.487	15.180 ***
Within 9 years	8,408	77.946	142.091	-3.052	3.279	-7.349	2.604 ***	-29.806	13.411 **	-36.845	12.730 ***
Within 10 years	8,407	69.179	136.326	-3.960	3.100	-6.197	2.245 ***	-24.986	11.077 **	-31.088	11.047 ***
<i>Panel B: Annual earnings (in 1,000 DKK)</i>											
Within 1 year	8,409	138.755	157.374	-2.389	2.215	-3.048	2.448	-3.813	14.350	-15.282	11.832
Within 2 years	8,409	136.581	158.471	-1.054	2.735	-2.710	2.611	-1.092	14.820	-13.588	12.708
Within 3 years	8,409	130.743	159.238	-0.624	3.285	-5.211	2.581 **	-13.604	14.176	-26.127	12.752 **
Within 4 years	8,409	126.535	158.446	-3.919	3.481	-9.253	2.921 ***	-33.420	16.111 **	-46.394	14.300 ***
Within 5 years	8,409	128.956	163.540	-3.895	3.301	-6.113	2.896 **	-17.671	15.689	-30.649	14.152 **
Within 6 years	8,409	133.934	168.489	-6.009	3.944	-6.912	3.130 **	-21.411	16.994	-34.654	15.006 **
Within 7 years	8,409	138.052	172.923	-6.633	4.223	-4.443	3.142	-7.794	17.565	-22.277	15.216
Within 8 years	8,409	135.594	175.152	-3.074	3.648	0.600	2.883	16.922	16.655	3.007	14.263
Within 9 years	8,409	123.820	171.656	-2.197	3.509	-0.154	2.530	12.404	14.762	-0.773	12.454
Within 10 years	8,409	110.753	165.789	-6.320	3.336 *	-5.861	2.856 **	-17.334	15.356	-29.387	13.787 **
Strata FE				YES		YES		YES		YES	
Additional Cov.				YES		YES		NO		YES	

*Notes:* In this table, I report: Ordinary Least Squares (OLS) estimates derived from the regression of the outcome variable on a dummy variable indicating whether or not individual  $i$  received a custodial sentence (equation 2); Reduced-Form (RF) estimates derived from the regression of my outcome variable on my instrument (equation 3); finally, I report the Instrumental Variables (IV) estimates obtained by instrumenting the dummy variable indicating whether or not individual  $i$  received a custodial sentence by my instrument. For each category of estimates, I report estimates obtained when a trend, a dummy variable indicative of whether or not the crime was committed in the 12 month-period preceding the entering into force of the reform, month of crime and district court fixed effects, and the whole conditioning set are added to the estimated equation (columns 1, 2, and 4). For IV estimates, I also report the results obtained when the conditioning set is not added to the estimated equation (column 3). Standard errors are clustered at the district court and individual levels. Significance levels: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .



## Appendix

**Table A.1: Legislation on drunk-driving crimes following the 1998 reform**

BAC (mg/l)		1st offense		2nd offense		3rd offense		4th offense		5th offense	6th offense	7th offense	8th offense
		DUI	Aggr. DUI	DUI	Aggr. DUI	DUI	Aggr. DUI	DUI	Aggr. DUI				
(0.50;0.80]	Fine	4.000 kr.	5.000 kr.	5.000 kr.	8.000 kr.	-	-	-	-				
	Prison	-	-	-	-	10 days	14 days	14 days	20 days				
	License susp.	-	Cond.	-	Cond.	-	Cond.	-	Cond.				
(0.80;1.20]	Fine	1 month's net pay	1 month's net pay	-	-	-	-	-	-				
	Prison	-	-	10 days	14 days	min. 30 days	min. 40 days	min. 60 days	min. 3 months	min. 4 months	min. 6 months		
	License susp.	Conditional	1 year	3 years	3 years	10 years*	Life suspension		Life suspension				
(1.20;1.50]	Fine	1 month's net pay	1 month's net pay	-	-	-	-	-	-				
	Prison	-	-	14 days	20 days	min. 30 days	min. 40 days	min. 60 days	min. 3 months	min. 4 months	min. 6 months		
	License susp.	1 year	2 years	3 years	3 years	10 years*	Life suspension		Life suspension				
(1.50;2.00]	Fine	1 month's net pay	-	-	-	-	-	-	-				
	Prison	-	14 days	14 days	30 days	min. 30 days	min. 40 days	min. 60 days	min. 3 months	min. 4 months	min. 6 months		
	License susp.	2 years	2.5 years	5 years	5 years	10 years*	Life suspension		Life suspension				
(2.00;2.50]	Fine	-	-	-	-	-	-	-	-				
	Prison	14 days	20 days	20 days	30 days	min. 30 days	min. 40 days	min. 60 days	min. 3 months	min. 4 months	min. 6 months		
	License susp.	2.5 years	3 years	5 years	5 years	10 years*	Life suspension		Life suspension				
>2.50	Fine	-	-	-	-	-	-	-	-				
	Prison	20 days	30 days	30 days	40 days	min. 30 days	min. 40 days	min. 60 days	min. 3 months	min. 4 months	min. 6 months		
	License susp.	2.5 years	3 years	5 years	5 years	10 years*	Life suspension		Life suspension				

\* Aggravating circumstances can lead to life suspension

**Table A.2: Additional information on the variables used**

<b>Conditioning set</b>	<b>Description</b> <i>N.B.: Names in italics refer to variables made available by Statistics Denmark. More information can be found on each of these variables on their website: <a href="https://www.dst.dk">https://www.dst.dk</a></i>
Offender's number of drunk-driving convictions in the 5 years preceding the crime	Computed from the date of crime. Top-coded at the 99 <sup>th</sup> percentile.  Variable source: we calculated this information myself using the exhaustiveness of the registers. Information on the date of the crime was retrieved from the <i>SIG_GERIDTO</i> variable.
Offender's number of other road traffic convictions in the 5 years preceding the crime	Computed from the date of crime. Top-coded at the 99 <sup>th</sup> percentile.  Variable source: we calculated this information myself using the exhaustiveness of the registers. Information on the date of the crime was retrieved from the <i>SIG_GERIDTO</i> variable.
Offender's number of non-road traffic convictions prior to the 5-year period preceding the crime	Computed from the date of crime. Top-coded at the 99 <sup>th</sup> percentile.  Variable source: we calculated this information myself using the exhaustiveness of the registers. Information on the date of the crime was retrieved from the <i>SIG_GERIDTO</i> variable.
Whether or not the offender was a juvenile at the time of the crime	Computed using information on the date of birth of the offender, as well as on the date of the crime.  Variable source: information on the offenders' date of birth was retrieved from the <i>FOED_DAG</i> variable and information on the date of the crime from the <i>SIG_GERIDTO</i> variable.
Offender's gender	Variable source: information on the gender of the offender was retrieved from the <i>AFG_KOEN</i> variable.
Offender's age at the time of the trial	Variable source: information on the gender of the offender was retrieved from the <i>AFG_AFGALD</i> variable.
Offender's immigration status	Dummy variables indicative of the following four groups of individuals: <ul style="list-style-type: none"> <li>- Immigrants</li> <li>- Descendants of immigrants</li> <li>- Unknown status</li> <li>- Rest of the population</li> </ul>

	<p>Variable source: information on the immigration status of the offender was retrieved from the <i>IE_TYPE</i> variable.</p>
Offender's marital status	<p>Dummy variables indicative of the following five groups of individuals:</p> <ul style="list-style-type: none"> <li>- Single</li> <li>- In a partnership</li> <li>- Separated</li> <li>- Widow</li> <li>- Unknown status</li> </ul> <p>Measured at the end of the year preceding the offender's crime.</p> <p>Variable source: information on the marital status of the offender was retrieved from the <i>CIVST</i> variable.</p>
Offender's highest educational achievement	<p>Dummy variables indicative of the following four groups of individuals:</p> <ul style="list-style-type: none"> <li>- Primary education</li> <li>- Secondary education</li> <li>- Higher education</li> <li>- Unknown highest educational achievements</li> </ul> <p>Measured at the end of the year preceding the offender's crime.</p> <p>Variable source: information on the education status of the offender was retrieved from the <i>HFFSP2</i> variable.</p>
Offender's earnings	<p>Annual earnings before tax and any other social contributions. Top-coded each year at the 99<sup>th</sup> percentile. Missing values were given the value 0.</p> <p>Measured at the end of the year preceding the offender's crime.</p> <p>Variable source: information on the offender's earnings was retrieved from the <i>LOENMV</i> variable.</p>
Offender's employment status	<p>Dummy variables indicative of the following six groups of individuals:</p> <ul style="list-style-type: none"> <li>- Full-time employment (&gt;29 hours per week)</li> <li>- Part-time employment (15-29 hours per week)</li> <li>- Limited employment (&lt;15 hours per week)</li> <li>- Employment with partial unemployment throughout the year</li> <li>- Unknown employment status</li> <li>- No employment</li> </ul>

	<p>Measured at the end of the year preceding the offender's crime.</p> <p>Variable source: information on the employment status of the offender was retrieved from the <i>JOBKAT</i> variable.</p>
--	--

<b>Outcome variables</b>	<b>Description</b> <i>N.B.: Names in italics refer to variables made available by Statistics Denmark. More information can be found on each of these variables on their website: <a href="https://www.dst.dk">https://www.dst.dk</a></i>
Offender's number of convictions following their trial	<p>Computed from the date of trial. We exclude from the computation of this variable convictions related to crimes committed prior to the trial of interest or convictions registered under the same criminal case identifier.</p> <p>Top-coded at the 99<sup>th</sup> percentile.</p> <p>Variable source: we calculated this information myself using the exhaustiveness of the registers. Information on the date of each crime was retrieved from the <i>SIG_GERIDTO</i> variable.</p>
Whether or not an offender was convicted following their trial	Computed from the above variable.
Offender's earnings	<p>Annual salary before tax and any other social contributions.</p> <p>Top-coded each year at the 99<sup>th</sup> percentile. Missing values were given the value 0.</p> <p>Measured at the end of each year, following the offender's trial.</p> <p>Variable source: information on the offender's earnings was retrieved from the <i>LOENMV</i> variable.</p>
Offender's number of days worked	<p>Annual number of days worked.</p> <p>Top-coded each year at the 99<sup>th</sup> percentile. Missing values were given the value 0.</p> <p>Measured at the end of each year, following the offender's trial.</p> <p>Variable source: information on the offender's earnings was retrieved from the <i>ANSDAGE</i> variable.</p>

**Table A.3: Reform and custodial sentence by subgroups**

Subgroups		Number of observations	Mean	Instrument Coeff. & s.e.
DD priors	<i>No prior conviction</i>	5,516	0.493	0.182*** (0.011)
	<i>Prior conviction(s)</i>	2,893	0.728	0.230*** (0.028)
Other priors	<i>No prior conviction</i>	4,149	0.562	0.162*** (0.020)
	<i>Prior conviction(s)</i>	4,260	0.585	0.234*** (0.026)
Age	<i>Below 25</i>	1,000	0.455	0.147*** (0.011)
	<i>25 and above</i>	7,409	0.590	0.207*** (0.000)
Gender	<i>Female</i>	707	0.636	0.233*** (0.010)
	<i>Male</i>	7,702	0.568	0.197*** (0.063)
Employment status	<i>Inactive</i>	2,058	0.573	0.176*** (0.014)
	<i>Unemployed</i>	912	0.607	0.225*** (0.013)
	<i>Employed</i>	5,368	0.569	0.206*** (0.013)
Education	<i>Higher education</i>	3,678	0.591	0.228*** (0.012)
	<i>Lower education</i>	4,434	0.561	0.179*** (0.015)

*Notes:* In this table, I estimate the first stage equation for various subgroups of the sample. More specifically, a dummy variable indicative of whether or not a defendant received a custodial sentence is regressed on my instrument, a trend, a dummy variable indicating whether the crime was committed in the 12-month period preceding the entering into force of the reform, month of crime and district court fixed effects, as well as the whole conditioning set. Standard errors are clustered at the district court and individual levels. Significance levels: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table A.4: Reform and other trial outcomes**

	(1)			(2)			<i>Sample</i>			(4)	
	DD crimes tried 0-6 months before the reform			DD crimes tried 0-6 months after the reform			DD crimes committed 0-24 months before the reform			Correlation between trial outcomes & instrument	
	#Obs.	Mean	S.d.	#Obs.	Diff.	S.e.	#Obs.	Mean	S.d.	Diff.	S.e.
Prison	1,754	0.668	0.471	1,943	0.016	0.009	8,409	0.733	0.443	0.001	0.009
Prison, uncond.	1,754	0.661	0.474	1,943	-0.508	0.010 ***	8,409	0.574	0.495	0.200	0.011 ***
Prison, uncond. (length in days)	1,754	16.558	23.675	1,943	-10.439	0.451 ***	8,409	15.042	25.390	3.764	0.558 ***
Imprisoned	1,754	0.305	0.461	1,943	-0.168	0.007 ***	8,409	0.346	0.476	0.055	0.009 ***
Prison, cond.	1,754	0.008	0.089	1,943	0.506	0.009 ***	8,409	0.155	0.362	-0.193	0.009 ***
Prison, cond. (length in days)	1,754	0.392	5.798	1,943	10.865	0.278 ***	8,409	3.575	10.317	-4.066	0.260 ***
Community work	1,754	0.000	0.000	1,943	0.290	0.006 ***	8,409	0.085	0.278	-0.108	0.005 ***
Fine	1,754	0.327	0.469	1,943	0.511	0.008 ***	8,409	0.420	0.494	-0.200	0.011 ***
Fine amount	1,754	1988.427	3464.473	1,943	2692.778	60.081 ***	8,409	2429.665	3541.650	-1120.090	66.350 ***
Driv. lic. suspended	1,754	0.290	0.454	1,943	0.000	0.008	8,409	0.309	0.462	-0.007	0.010
Driv. lic. suspended (length in days)	1,754	1.741	2.724	1,943	0.001	0.045	8,409	1.854	2.773	-0.041	0.058
Appeal	1,754	0.010	0.098	1,943	0.028	0.003 ***	8,409	0.020	0.140	-0.001	0.003

*Notes:* In columns (1) and (2), I describe the outcomes of drink-driving cases tried in the 6-month period before the entering into force of the reform and compare these with the outcomes of drink-driving cases tried in the following 6-month period. In order to do the latter, I regress trial outcomes on a constant, a dummy variable taking the value 1 when a case was tried after the reform and district court fixed effects. The coefficients associated with the dummy variable are reported together with their standard errors. In columns (3) and (4), I describe the outcomes of drink-driving cases included in my sample and investigate the extent to which their outcome is correlated with my instrument. In order to do so, I regress trial outcomes on the instrument, a time trend, a dummy variable indicative of whether or not the crime was committed in the 12-month period preceding the entering into force of the reform, and month of crime and district court fixed effects. The coefficients associated with the instrument are reported together with their standard errors. Standard errors are clustered at the district court and individual levels. Significance levels: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**D.7 How does transportation shape intrametropolitan growth? An answer from the Regional Express Rail**

**ORIGINAL ARTICLE**

# How does transportation shape intrametropolitan growth? An answer from the Regional Express Rail

Miquel-Àngel Garcia-López<sup>1</sup> | Camille Hémet<sup>2</sup> |Elisabet Viladecans-Marsal<sup>3</sup>

<sup>1</sup>Universitat Autònoma de Barcelona and Institut d'Economia de Barcelona, Departament of Applied Economics, Edifici B, Facultat d'Economia i Empresa, 08193 Cerdanyola del Vallès, Spain

<sup>2</sup>Paris School of Economics and Institut d'Economia de Barcelona and CEPR, 48 Boulevard Jourdan, 75014 Paris, France

<sup>3</sup>Universitat de Barcelona and Institut d'Economia de Barcelona and CEPR, Department of Economics, John M Keynes 1-11, 08034 Barcelona, Spain

**Correspondence**

Miquel-Àngel Garcia-López, Universitat Autònoma de Barcelona and Institut d'Economia de Barcelona, Departament of Applied Economics, Edifici B, Facultat d'Economia i Empresa, 08193 Cerdanyola del Vallès, Spain.  
Email: miquelangel.garcia@uab.cat

**Abstract**

This paper analyzes the influence of transportation infrastructure, and in particular of the Regional Express Rail (RER), on employment and population growth in the Paris metropolitan area between 1968 and 2010. In order to make proper causal inference, we rely on historical instruments and control for all other transportation modes that could be complement or substitute to the RER. Our results show that proximity to an RER station increases employment and population density and, in particular, employment and population growth. The latter effects are higher in municipalities located near RER stations and close to employment (sub)centers. They are also found to be particularly strong for jobs in the service sector, for factory workers, and for highly educated population. We find no impact of the RER expansion on employment growth during the first part of the period, while the effect on population growth appears earlier but declines over time.

**KEYWORDS**

transportation, urban growth, urban spatial structure

## 1 | INTRODUCTION

During recent decades, the city of Paris and its metropolitan area have undergone major demographic and socioeconomic changes. Specifically, between 1968 and 2010, the Paris metropolitan area strengthened its position both as the most populated area (around 12 million inhabitants) and as the largest economic region in the country (with one fifth of total employment). However, this growth has not been homogeneously distributed throughout the metropolitan area. In fact, even though the levels of employment and population of the whole area respectively grew by 32 percent and 27 percent over this period, the central business district (CBD) has seen employment fall by 7 percent and population by 13 percent. As a result of these changes, the share of total metropolitan area employment located in the CBD fell from 45 percent in 1968 to 32 percent in 2010. The same pattern is observed for the population of the CBD, which today represents around 19 percent of the whole area, compared to 28 percent at the end of the 1960s. All in all, these trends indicate that the Paris metropolitan area has undergone a marked process of

suburbanization, accompanied by the emergence of employment subcenters (areas of high employment density outside the CBD).

There is a long tradition in the literature of studies seeking to explain the determinants of city structure and city growth. An important amount of research has been devoted to understanding why some cities are more successful and grow more rapidly than others, and how the urban structure changes, leading to a variety of responses: while some authors emphasize the importance of human capital and skills (Moretti, 2004; Rosenthal & Strange, 2008, are good examples), others focus on the role of the weather (Glaeser, Kolko, & Saiz, 2001) or of the availability of consumer amenities (Carlini & Saiz, 2008) as attractors for population. Ultimately, the literature agrees that a city's density largely accounts for its capacity to be productive and to attract better firms and workers (Combes, Duranton, Gobillon, Puga, & Roux, 2012). For urban economists, agglomeration economies are therefore considered an important source of city growth. What is also clear is that the advantages provided by agglomeration economies increase with a reduction in the transportation costs for goods and people (Glaeser & Gottlieb, 2009). Yet, even though transportation seems to be a key element for employment and population distribution and growth, its impact on urban growth has long been ignored in the empirical literature (see the pioneering work by Duranton & Turner, 2012).

The main objective of this paper is thus to contribute to this relatively recent branch of the literature by analyzing the spatial influence on employment and population growth of a major improvement in transportation infrastructure. More precisely, our study relies on the case of the Paris metropolitan area, where the transportation network has been substantially improved with the introduction of a new suburban train, the *Réseau Express Régional* (RER henceforth). Since its inauguration in 1975, the total length of the RER network has increased by around 550 km, and now operates 257 stations connecting more than 170 municipalities. Yet, this is not the only change in the transportation infrastructure that the region has undergone over the past 40 years. Likewise, the metro and tramway networks have also been expanded throughout the area. Most notably, the area's main road system (highways) has been extended by 600 km, while the number of ramps and accessibility to many other municipalities have been increased. Thus, although we focus our attention on the expansion of the RER (1975–2010) and its effects on the location of new jobs and inhabitants, we need to take into account all the other transportation modes that might complement or substitute the RER system. We therefore make considerable efforts to control for the changes in all other transportation modes of the area in order to estimate unbiased estimate an unbiased effect of the RER. We first study changes in the urban spatial structure of the Paris metropolitan area between 1968 and 2010, as well as the role of the CBD's and employment subcenters' location. We then turn to analyzing the spatial influence of transportation infrastructure on the 2010 intrametropolitan distribution of employment and population. We finally estimate whether transportation fostered employment and population growth during this period.

Doing so, this paper makes three important contributions to the literature. First, we analyze the impact of RER improvements on employment and population growth for each of the 1,300 municipalities in the Paris metropolitan area, which underwent important improvements in its transportation system in recent decades. Second, to the best of our knowledge, this is the first work to undertake an analysis of the causal effects of improvements to an infrastructure system on city growth controlling for all other possible modes of transportation (railroads, metro, tramways, and highways). Third, using historical instruments allows us to solve the endogeneity problem that is common in this literature.

Our work builds on seminal theoretical contributions which help frame the way transportation infrastructures should affect population and employment growth: the classical monocentric city model developed by Alonso (1964), Mills (1967), and Muth (1969) shows that transportation (accessibility) is the main factor that determines urban land use (Duranton & Puga, 2015). Anas and Moses (1979) and Baum-Snow (2007) extend this model by considering two competing transportation infrastructures: a classical transportation infrastructure based on a dense network of radial streets and a high-speed transportation infrastructure based on sparse radial corridors (Anas & Moses, 1979; Baum-Snow, 2007). Depending on the cost of alternative transportation modes, the authors find that population and employment spread out along the sparse corridors, increasing surrounding land rents and densities. This paper can thus be viewed as an empirical test of these theoretical predictions regarding the role of transportation infrastructures. While we focus on the effect on the spatial distribution and growth of population and employment across municipalities of a large metropolitan area, other empirical studies analyze the effects of railroads on a wide range of

alternative outcomes. A great deal of papers study for instance the impact of new railroads and stations on the housing market and the property values of the affected area. Baum-Snow and Kahn (2000), Baum-Snow and Kahn (2005), Bowes and Ihlanfeldt (2001), Gibbons and Machin (2005), and Ryan (2005) are good examples, but are mostly applied to U.S. cities. More recent studies, such as Diao, Zhu, and Zhu (2016) and Li, Yang, Qin, and Chonabayashi (2016), focus instead on Chinese cities and analyze the impact of the high speed train on the housing markets. On the same topic, other recent approaches explore the incidence of railroads on land use (Hurst & West, 2014; Schuetz, 2015). Closer to our paper, Chatman and Noland (2014) study the impact on employment density and productivity in U.S. central cities, and Kotavaara, Antikainen, and Rusanen (2011) focus on population growth in Finish municipalities, but do not take stand on the spatial structure of these cities. Some recent research departs from these traditional topics in urban economics, and has started looking into the effect of railroad on largely unexplored aspects, such as air pollution (Chen & Whalley, 2012) or cities' gentrification (Grube-Cavers & Patterson, 2015). The evidence of these analyses is quite diverse, but a majority of papers point to a positive or neutral effect of these investments on the different analyzed outcomes. Our paper distinguishes itself from this large literature by analyzing the incidence of railroads investments on the internal spatial structure of cities, which remains largely unexplored, to the best of our knowledge.

Our study is also related to recent empirical studies that have examined other aspects of transportation infrastructure. Sharing our intrametropolitan approach, some papers focus on the effect of rail or road transportation on the suburbanization process, both in the United States (Baum-Snow, 2007) and in Europe (Garcia-López, Holl, & Viladecans-Marsal, 2015; Garcia-López, Pasidis, & Viladecans-Marsal, 2015), while other studies, conducted at the county level, consider alternative outcomes, such as workers' earnings (Michaels, 2008) or employment growth (Jiwattanakulpaisarn, Noland, Graham, & Polak, 2009). Duranton and Turner (2011) and Hsu and Zhang (2014) provide intermetropolitan evidence for the effect of highway improvements on congestion in the United States and Japan, respectively. In the development economics literature as well, some recent papers analyze the effect of infrastructures on various economic outcomes, in China (Banerjee, Duflo, & Qian, 2012; Faber, 2014) or India (Donaldson, 2015).

Analyzing the impact of infrastructure improvements on city growth typically involves two problems of inference. First, all types of infrastructure take time to be built and their effects on city growth are not immediate. This problem can be solved by using long differences for both employment or population and infrastructure changes. Second, the location of new infrastructures is not random. Rather, it is most likely endogenous to employment and population growth: planners may decide to improve the connection of deprived areas in order to boost their economic activity or attract population, or they may on the contrary connect areas that are expected to grow, anticipating on future needs for transportation. It means that any naive estimate of the effect of the distance to a transportation infrastructure on population or employment growth will be biased, which is one of the main issues in the literature. Only recently have a few papers proposed various inference strategies to address this problem. A first category of solution consists in using exogenous deviations from initial transportation plans (see, for instance, Baum-Snow, 2007; Michaels, 2008). Special mention needs to be made to Mayer and Trévien (2015) who also focus on the Paris region. They exploit the deviation from the initial investment plan for the RER network resulting from budgetary and technical constraints to evaluate the impact of the opening and of the progressive extension of the RER between 1975 and 1990 on employment and population at the municipal level. An alternative identification strategy has been proposed by Jiwattanakulpaisarn et al. (2009), who use the lagged levels of highway lane-mile density as instruments for highway infrastructure investments. Finally, the use of historical instruments is another popular strategy to tackle this causality issue. For instance, Duranton and Turner (2012) instrument road infrastructure using the U.S. railway network at the end of the 19th century and the routes taken by major expeditions of the United States between 1518 and 1850; Hsu and Zhang (2014) rely on the 1890 railway network plan and the planned national express way extension as exogenous sources of variation of highway location in Japan; and Garcia-López et al. (2015) use the Roman roads and the 1760 Postal routes as instruments for Spanish highways. Our empirical strategy follows this approach, as we rely on two historical instruments, the 1870 railways and the Roman roads, as a source of exogenous variation of current infrastructure' location.

Our results show that the RER network influences the location of employment and population, even after controlling for other modes of transportation. Getting 1 km closer to an RER station is found to increase employment and

population density by around 5 percent and 4 percent, respectively. Furthermore, a dynamic analysis reveals that improving the RER network significantly increases municipal employment and population growth: for each kilometer closer to an RER station, employment increases by 2 percent and population by 1 percent. Although this impact seems limited, it is considerably reinforced once we introduce heterogeneity in the analysis: (1) For municipalities located less than 1 km from an RER station, each kilometer closer to an RER station increases employment and population growth by 8 percent and 6 percent, respectively; (2) For noncentral municipalities located less than 5 km from an employment (sub)center, each kilometer closer to an RER station increases employment and population growth by 12 percent and 9 percent, respectively; (3) The effects on employment growth are higher for services (3 percent) and factory jobs (3 percent), the effects on population growth are higher for inhabitants with high school (3 percent) and university (2 percent) degree; (4) We find no impact of the RER expansion on employment growth during the first part of the covered period, while the impact on population growth was sizeable much earlier but declined over time.

The remainder of the paper is organized as follows. Section 2 first describes the changes in the urban spatial structure in the Paris metropolitan area and highlights the suburbanization process of the area. It then turns to presenting the main changes in the different transportation infrastructures in the Paris metropolitan area and their influence on the intrametropolitan location of employment and population. Section 3 presents our main results, and Section 4 concludes.

## 2 | URBAN SPATIAL STRUCTURE AND TRANSPORTATION IN THE PARIS METROPOLITAN AREA

### 2.1 | Urban spatial structure and suburbanization process

The *Paris metropolitan area* is one of the 22 administrative regions in continental France, known as *Ile de France*, to which the city of Paris belongs. It is divided into eight *départements* (administrative subregions) and 1,300 municipalities. Note that the city of Paris has been a *département* of its own since 1968, and is divided into 20 municipalities called *arrondissements*. The municipality is the unit of analysis of this paper. It is actually the smallest administrative division that we can use, since smaller divisions were not introduced in the French census before 1999. This is however a reasonable unit of analysis given our research agenda as French municipalities are particularly small<sup>1</sup>: in *Ile de France*, the average municipal surface is 9.3 km<sup>2</sup>, and the median is 7.6 km<sup>2</sup>. The metropolitan area of Paris is the densest and the most populated region in France, with 986.7 inhabitants per square kilometer in 2011 for a total of 11,852,851 inhabitants. It is also the main employment center in the country: with a total of 5,660,253 jobs in 2011, it accounted for more than a fifth of total employment in continental France. Among them, 0.2 percent work in the agricultural sector, 5.2 percent in the construction sector, 8.4 percent in industry, and the remaining 86.2 percent in the tertiary sector (trade, services, public administration, education, among others).

With its 20 *arrondissements* expanding over 105.4 km<sup>2</sup> for a density of 21,347 inhabitants per square kilometer, the city of Paris constitutes the CBD of the Paris metropolitan area. In 2011, about 2,250,000 inhabitants lived in Paris, corresponding to 19 percent of the metropolitan area's total population. The CBD also accounted for 32 percent of the metropolitan area's employment, with about 1,800,000 jobs, concentrated in the tertiary sector: 67.9 percent in trade, transportation, and services and the remaining 24.4 percent in public administration, education, health, and social services. We can notice from these figures that tertiary sector jobs are overrepresented in the CBD: 92.3 percent of all jobs in Paris compared to 86.2 percent of all jobs in the whole metropolitan area.

As most large agglomerations, the Paris metropolitan area also includes several employment subcenters. We identify them using the method first developed by McDonald and Prather (1994) and improved by McMillen (2001). The general idea is to estimate densities following a monocentric spatial pattern. The predicted densities obtained are then

<sup>1</sup> Mainland France comprises more than 36,500 municipalities.

subtracted from the corresponding real densities. From these residuals, those that are positive and statistically significant are selected and defined as subcenters.<sup>2</sup> This method, which is described more thoroughly in Appendix A, enables us to identify 21 employment subcenters made of 88 municipalities in 1968 and 34 subcenters including 89 municipalities in 2010.

After identifying the main characteristics of the urban spatial structure in Paris metropolitan area, we now study its temporal and spatial trends: we characterize the importance of the CBD, the subcenters, and the other municipalities, before looking at the spatial influence of the CBD and the subcenters on the location of firms and households.

All information about population and employment used in this paper comes from various censuses provided by the French statistical institute, the INSEE.<sup>3</sup> Each census wave provides us with the number of individuals living and working in each municipality, and enables us to determine their socioeconomic and demographic characteristics (e.g., level of education, socioeconomic category of the job, type of occupation, age, gender, nationality, birth country, marital status, household size). Because our study is based on the 1968–2010 period, during which the railroad network underwent significant improvements (further details on this topic are given in Section 2.2), we therefore rely on the 1968, 1975, 1982, 1990, 1999, and 2010 waves of the census.

Table 1 reports the number of jobs (Panel A) and inhabitants (Panel B) in the CBD, the identified subcenters and the remaining municipalities in 1968 and 2010. We alternatively refer to subcenters identified in 1968 (first three columns) or in 2010 (last three columns). The total numbers in the bottom line of each panel reveal that the Paris metropolitan area as a whole grew by about one third over the period, both in terms of employment (32.6 percent) and population (27.4 percent). Disaggregating these figures between CBD, subcenters, and other locations enables us to detect the suburbanization process experienced by the Paris metropolitan area since 1968. Indeed, we see that the number of jobs in the CBD decreased by 7.1 percent, while population fell by 13.4 percent, to the benefit of subcenters and other municipalities. This evolution reflects an absolute suburbanization process. We can also note that the CBD's share of total employment and population dropped, respectively, from 45.3 percent to 31.7 percent and from 28 percent to 19 percent, revealing that the CBD's decentralization process was also relative. Taking a closer look at the subcenters and comparing the 1968 situation of the subcenters identified in 1968 with the 2010 situation of those identified in 2010, we can observe that they gained in terms of employment, both in absolute and relative terms (from 33 percent to 35 percent), illustrating a process of absolute and relative employment centralization in the subcenters. On the other hand, subcenters lost in terms of population, both in absolute and relative terms (from 34 percent to 26 percent), to the benefit of other municipalities: the subcenters have themselves been undergoing a population suburbanization process toward smaller municipalities.

In order to get an idea of whether the CBD and the subcenters influence the intrametropolitan distribution of employment and population, we regress the 2010 employment and population densities (in log) on the distance to CBD and the distance to the nearest employment subcenter (where we alternatively use subcenters identified in 1968 and 2010), controlling for a vector of geographic characteristics (land area, altitude, ruggedness index, and elevation range). Municipal employment and population densities are found to be larger as the municipality gets closer to the CBD or to a subcenter (this trend is particularly marked for population density), revealing that both CBD and subcenters influence the spatial pattern of employment and population location. If we perform the same exercise considering population or employment density *growth* as the dependent variable, we observe that this growth was larger for municipalities closer

---

<sup>2</sup> The empirical literature has proposed alternative procedures to identify subcenters. Among them, the most used are those based on density and employment thresholds (Giuliano & Small, 1991; Giuliano, Redfearn, Agarwal, Li, & Zhuang, 2007; Muñiz, Garcia-López, & Galindo, 2008) and on employment and/or population density peaks (McDonald, 1987; McDonald & Prather, 1994; Sivitanidou, 1996; Craig & Ng, 2001; Muniz et al., 2003; Redfearn, 2007; Garcia-López, 2010; Craig, Kohlhase, & Perdue, 2016). Although they sometimes identify different sets of subcenters in the same city (for instance, Redfearn (2007) finds different subcenters in Los Angeles when using his own methodology than when relying on Giuliano and Small (1991) and McMillen (2001)'s procedures), there are examples in which different methods lead to similar sets of subcenters (see for instance Craig et al. (2016) in Houston when using McMillen (2001) and Redfearn (2007)'s strategies). In our case, we rely on McMillen (2001)'s method not only because it is consistent with theory but also because it is directly linked and based on (deviations from) the monocentric model, which is consistent with the historical spatial structure of the Paris metropolitan area.

<sup>3</sup> Between the early 1960s and the late 1990s, one census surveying all individuals living in France was conducted about every decade. Since 2004, the design and sampling methodology of the census has changed completely, and is now conducted annually over a fraction of the population, so that the census data labeled "year  $n$ " is in fact collected over five years ( $n - 2$  to  $n + 2$ ). More details on this new sampling methodology can be found in English on the INSEE webpage.

**TABLE 1** Employment and population in metropolitan Paris, 1968–2010

	1968 Subcenters		2010 Subcenters		1968–2010	
	1968	2010	1968	2010	1968	2010
Panel A: Employment						
Paris	1,935,716 (45.26%)	1,797,678 (31.71%)	1,935,716 (45.26%)	1,797,678 (31.71%)	–138,038 (–7.1%)	–138,038 (–7.1%)
Subcenters	1,419,072 (33.18%)	1,762,894 (31.10%)	1,132,124 (26.47%)	1,978,722 (34.91%)	846,598 (74.8%)	846,598 (74.8%)
Noncentral municipalities	921,992 (21.56%)	2,108,330 (37.19%)	1,208,940 (28.27%)	1,892,502 (33.38%)	683,562 (56.5%)	683,562 (56.5%)
Total	4,276,780	5,668,902	4,276,780	5,668,902	1,392,122 (32.6%)	1,392,122 (32.6%)
Panel B: Population						
Paris	2,590,771 (28.01%)	2,243,833 (19.04%)	2,590,771 (28.01%)	2,243,833 (19.04%)	–346,938 (–13.4%)	–346,938 (–13.4%)
Subcenters	3,153,224 (34.10%)	3,472,991 (29.46%)	2,370,046 (25.63%)	3,103,007 (26.33%)	732,961 (30.9%)	732,961 (30.9%)
Noncentral municipalities	3,504,637 (37.89%)	6,069,410 (51.50%)	4,287,815 (46.36%)	6,439,394 (54.63%)	2,151,579 (50.2%)	2,151,579 (50.2%)
Total	9,248,632	11,786,234	9,248,632	11,786,234	2,537,602 (27.4%)	2,537,602 (27.4%)

Notes: Metropolitan shares and growth rates in parentheses.

to the CBD and subcenters, suggesting an increasing influence of both CBD and subcenters on the location of jobs and residences despite the suburbanization process.<sup>4</sup>

## 2.2 | Transportation and the location of employment and population in 2010

The transportation infrastructure of the Paris metropolitan area today is based on both a railroad network and a main road system. There are four types of railway networks. First, a suburban train (train henceforth) that connects Paris to the suburbs, including some of the most remote parts of the region. This network was initiated during the first half of the nineteenth century, and has been continuously expanded since then. An important modernization wave took place in the 1960s, with all steam trains being replaced by electric trains. Panel A of Table B.1 in Appendix B shows the most recent evolution of the train network. This network, composed of five lines with a total length of 788 km and with 231 stations located in 196 municipalities in 2010, has undergone a slight reduction in the past 40 years.

The Paris region is also endowed with a regional express network, the RER (*Réseau Express Régional*), which started operating during the second half of the 1970s. Figure B.1 in Appendix B shows the evolution of the RER network between 1975 and 2010. Like the train, the RER connects Paris to the suburbs, but for a shorter maximum distance: the furthest RER stations are located about 30 km away from Paris. Most of the RER lines follow the train lines and were designed to improve the former network. An important distinction between the train and RER networks is that the latter has connections within Paris. The RER thus enables passengers to commute from one part of the Paris metropolitan area to another, going through Paris, but without having to switch to another train or metro to cross the city. This represents a clear improvement to regional transit overall. Between 1975 and 2010, the RER network increased its number of lines from one to five, its total length from 39 to 587 km, its number of stations from 22 to 243, and the number of municipalities having an RER station grew from 16 to 167 (see Appendix B Table B.1 Panel B). We refer to Mayer and Trévien (2015) for a detailed history of the rail network in the Paris metropolitan area, and for a thorough explanation of the differences between the two regional train networks.

In addition to these regional railroad networks, Paris is endowed with a very dense subway system (*métro* henceforth), which was opened in 1900 and mainly connects areas within Paris. Between 1968 and 2010, the *métro* network was further expanded with the addition of two new lines that increased its length by 44 km, and 34 new stations were added connecting 13 new municipalities (Appendix B Table B.1 Panel C). Today, a few *métro* stations extend beyond Paris, but they remain within a very limited range (contiguous municipalities mostly). Finally, the Paris metropolitan area also enjoys a tramway network, which is much more recent: the first segments started operating in the beginning of the 1990s, and the network is still expanding. This network is mostly located at the fringe of Paris, with some segments running in the first ring of municipalities around Paris. Note that while the main regional trains have a radial structure, linking Paris to the suburbs, the tramway is much more circular, the various lines forming a circle around the CBD and along its borders. In 2010, this network was based on four lines with a total length of 40 km, with 70 stations connecting 19 municipalities (Appendix B Table B.1 Panel D).

In the case of the main road system, we focus on the highway network (and include some other main roads). Although France's first highway projects date from the 1920s and the 1930s, the real expansion of the French network took place during the second half of the 20th century. In the Paris metropolitan area (Appendix B Table B.1 Panel E), the number of highways increased from 11 to 41 between 1968 and 2010, expanding the network from 229 km with 46 ramps in 40 municipalities to 821 km with 168 ramps in 133 municipalities.

We now want to investigate the spatial influence of transportation on the intrametropolitan distribution of employment and population in the Paris metropolitan area in 2010. As mentioned in the Introduction, this analysis is likely to be flawed by the identification issue that is pervasive in the literature: transportation and its improvements are not decided and located randomly. On the contrary, they are endogenous to employment and/or population growth: planners may decide to improve the connection of deprived areas in

order to boost their economic activity or attract population, or they may on the contrary connect areas that are expected to grow, anticipating on future needs for transportation. Following recent literature (Duranton & Turner, 2012; Garcia-López, 2012; Garcia-López, Pasidis, & Viladecans-Marsal, 2015), we rely on historical railways and roads to instrument the location of modern transportation infrastructures in the Paris metropolitan area. More precisely, we use Roman roads and the 1870 railroads as alternative instrumental variables. We provide evidence of the validity in terms of exogeneity and relevance of these instruments in Appendix C.

Since our focus is on the RER, we regress the log of the 2010 employment and population densities on the distance to the nearest RER station in 2010, controlling for the distance to other types of transportation infrastructures and for the distance to the nearest 2010 employment center, where distances are alternatively included as such or in logarithm.<sup>5</sup> We also control for geographical characteristics (land area, altitude, ruggedness index and elevation range), as well as for historical dummy variables indicating (1) whether municipalities were Roman settlements, (2) whether they used to be major towns between the 10th and the 15th centuries, (3) whether they used to be major towns between the 16th and the 19th centuries, (4) whether they had a monastery built between the 12th and 16th centuries, and (5) whether they hosted important fairs between the 10th and the 16th centuries<sup>6</sup>:

$$\begin{aligned}
 2010 \ln(\text{density}) = & \delta_0 + \delta_1 \times 2010 f(\text{distance to RER station}) \\
 & + \delta_2 \times 2010 f(\text{distance to non-RER station or ramp}) \\
 & + \delta_3 \times f(\text{distance to the nearest 2010 employment center}) \\
 & + \sum_i (\delta_{4,i} \times \text{geography}_i) + \sum_i (\delta_{5,i} \times \text{history}_i). \tag{1}
 \end{aligned}$$

Because we either use direct measures of distances in kilometers or their logarithm, the  $f$  function stands alternatively for the identity function or the logarithm function, respectively. The coefficients  $\delta_1$  and  $\delta_2$  represent density gradients when we use direct distances, and elasticities when we use logged distances. They capture the extent to which density increases with proximity to the nearest RER station and to the nearest non-RER station or highway ramp, respectively. In order to address the endogeneity issue, we estimate this equation using a two-stage least square (TSLS) procedure, where Roman roads and 1870 railroads are used as instruments for the RER and the non-RER variables (see details in Appendix C).

Table 2 reports results from estimating Equation (1) for employment density (columns 1–5) and population density (columns 6–10) when using direct distances (Panel A) and logged distances (Panel B). In all cases, we find that transportation infrastructures do influence the location of employment and population: the estimated coefficients are always negative and significant. In particular, our results show that getting closer to an RER station by 1 km (or 10 percent if we read our results in terms of elasticities) increases employment density by around 6 percent and population density by 4 to 5 percent. Results also show higher (but less significant) effects for non-RER transportation: each additional kilometer closer to a non-RER station or ramp increases employment and population densities by 7 to 8 percent and 6 to 7 percent, respectively. In terms of elasticities, we can say that getting 10 percent closer to a non-RER infrastructure leads to a 5 to 10 percent rise in employment density and to a 4 to 9 percent rise in population density. In the rest of the paper, we focus on results controlling only for the group of non-RER infrastructures, including highways ramps (i.e., the specifications used in columns 5 and 10 of Table 2), since these non-RER coefficients are of the same order of magnitude (considering their standard errors), and do not affect the RER coefficient much.

<sup>5</sup> As Redfearn (2009) and McMillen and Redfearn (2010) highlight, it is difficult to choose functional forms in spatial data. As a result, we use the most common functional forms in the literature: direct and logged distances.

<sup>6</sup> These variables come from the Digital Atlas of Roman and Medieval Civilizations, with the exception of the major cities of the 16th to 19th centuries, which are identified in Bairoch (1988).

TABLE 2 Urban spatial structure and proximity to RER and other transportation, TSLS

Dependent variable:	2010 ln[Employment density]			2010 ln[Population density]						
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
<i>Panel A: Direct distances</i>										
2010 Distance to the nearest RER station	-0.065* (0.007)	-0.063* (0.007)	-0.063* (0.007)	-0.052* (0.009)	-0.055* (0.008)	-0.050* (0.005)	-0.050* (0.006)	-0.050* (0.005)	-0.041* (0.007)	-0.043* (0.006)
2010 Distance to the nearest commuter train		-0.074*** (0.039)				-0.061** (0.030)				
2010 Distance to the nearest non-RER station			-0.079*** (0.042)					-0.066** (0.032)		
2010 Distance to the nearest highway ramp				-0.072*** (0.038)					-0.060** (0.030)	
2010 Distance to the nearest non-RER stat/ramp					-0.084*** (0.044)					-0.069** (0.034)
First-stage statistic	226.48	38.46	34.00	23.68	54.16	226.48	38.46	34.00	23.68	54.16

(continues)

TABLE 2 (Continued)

Dependent variable:	2010 ln(Employment density)					2010 ln(Population density)				
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
	Panel B: Logged distances									
2010 ln(Dist to the nearest RER station)	-0.679* (0.067)	-0.620* (0.066)	-0.605* (0.067)	-0.542* (0.084)	-0.550* (0.075)	-0.550* (0.052)	-0.498* (0.051)	-0.485* (0.051)	-0.429* (0.069)	-0.436* (0.057)
2010 ln(Dist to the nearest commuter train)		-0.432* (0.150)				-0.380* (0.116)				
2010 ln(Dist to the nearest non-RER station)			-0.472* (0.161)					-0.415* (0.125)		
2010 ln(Dist to the nearest highway ramp)				-1.029** (0.410)	-				-0.905* (0.338)	
2010 ln(Dist to the nearest non-RER stat/ramp)					-0.536* (0.185)					-0.471* (0.145)
First-stage statistic	172.46	48.71	40.21	10.13	35.23	172.46	48.71	40.21	10.13	35.23
Distance to the nearest 2010 center	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geography	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
History	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Instrument:										
Distance to the nearest 1870 railroad line	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Distance to the nearest Roman road	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: One thousand three hundred observations for each regression. Geography variables are land area, altitude, index of terrain ruggedness, and elevation range. History variables are dummy variables for municipalities (1) that were Roman settlements (based on DARMC maps), (2) that were major towns between the 10th and the 15th centuries (based on DARMC maps), (3) that were major towns between the 16th and the 19th centuries (based on Bairoch, 1988), (4) with a monastery built between the 12th and 16th centuries (based on DARMC maps), and (5) that hosted important fairs between the 10th and the 16th centuries (based on DARMC maps). Robust standard errors are in parentheses. \*, \*\*, and \*\*\* indicates significant at 1, 5, and 10 percent level, respectively.

### 3 | THE EFFECT OF THE RER ON LOCAL GROWTH IN THE PARIS METROPOLITAN AREA

By investigating the effects of transportation's improvements on local growth in employment and population in the Paris metropolitan area, our paper brings several new insights to the related literature. First, it focuses on the intrametropolitan level, that is, on the municipalities that make up the Paris metropolitan area, while most previous studies are at the metropolitan level. Second, we study the effects on both employment and population growth, while previous studies focus on just one or the other. Finally, although our main interest is the effect of the RER, we also control for other modes of transportation, while most of the previous studies consider just one type of infrastructure.

In this section, we start by studying the RER effects on local growth for all 1,300 municipalities that make up Paris metropolitan area. We then explore whether the RER effects are heterogeneous across space by grouping municipalities according to their proximity to an RER station and to their proximity to employment (sub)centers. Afterwards, we study heterogeneous effects according to the type of employment (sectors and occupations) and the education level of population, before analyzing the temporal scope of the RER effects across time periods.

#### 3.1 | Average metropolitan effects

We begin by analyzing the impact of the RER and other transportation modes on local (municipal) growth, both in terms of employment and population. Since it might take years for firms and households to relocate in response to transportation improvements, we estimate traditional growth equations, in which the growth rate of the dependent variable between years  $t$  and  $t - 1$  is regressed on a set of explanatory variables measured in the initial year  $t - 1$ . Here, we focus on the 1968–2010 period and estimate the following equation:

$$\begin{aligned}
 1968 - 2010 \Delta \ln(\text{density}) = & \mu_0 + \mu_1 \times 2010 f(\text{distance to RER station}) \\
 & + \mu_2 \times 1968 f(\text{distance to non-RER stations and ramps}) \\
 & + \mu_3 \times 1968 \ln(\text{densities}) \\
 & + \mu_4 \times f(\text{distance to the nearest 1968 employment center}) \\
 & + \sum_i (\mu_{5,i} \times \text{geography}_i) + \sum_i (\mu_{6,i} \times \text{history}_i) \\
 & + \sum_i (\mu_{7,i} \times 1968 \text{ socioeconomy}_i). \tag{2}
 \end{aligned}$$

It is important to point out that, since there were no RER stations in 1968, our main explanatory variable is the 2010 distance to the nearest RER station (or its logarithm depending on specifications). On the other hand, since there were other railroads and highways in the initial year, we include the distance to the nearest access to a non-RER infrastructure (station or ramp) in 1968 (or its logarithm). We also control for characteristics related to the initial urban spatial structure of the Paris metropolitan area, that is, the 1968 employment and population densities, the (log) distance to the nearest 1968 employment center, and geography and history variables. We additionally control for the 1962 population size, and 1968 socioeconomic characteristics at the municipal level: unemployment rate; share of employment in manufacturing, in construction, and in services, used as proxies for economic specialization; share of executives and professional workers, to account for average income level; and share of population with university degree, as a proxy for the level of human capital. Here again, we run TSLS regressions using distance to the nearest 1870 railroad and distance to the nearest Roman road as instruments to correct for endogeneity (see Appendix C for further details).

Table 3 reports our main TSLS results for employment (columns 1–3) and for population (columns 4–6) when using direct distances (Panel A) and logged distances (Panel B). In order to understand the determinants of employment

**TABLE 3** The effect of RER on municipality growth, TSLS: Average metropolitan effects

Dependent variable:	1968–2010 $\Delta \ln(\text{Employment density})$		1968–2010 $\Delta \ln(\text{Population density})$			
	[1]	[2]	[3]	[4]	[5]	[6]
<i>Panel A: Direct distances</i>						
2010 Distance to RER station	-0.018* (0.005)		-0.023* (0.005)	-0.011* (0.003)		-0.014* (0.004)
1968 Distance to non-RER stat/ramp		-0.068* (0.023)	-0.089* (0.022)		-0.048* (0.015)	-0.061* (0.015)
1968 $\ln(\text{Employment density})$	-0.543* (0.078)	-0.574* (0.080)	-0.543* (0.076)	-0.011 (0.058)	-0.030 (0.058)	-0.011 (0.057)
1968 $\ln(\text{Population density})$	0.629* (0.089)	0.653* (0.075)	0.463* (0.088)	0.038 (0.063)	0.044 (0.050)	-0.075 (0.061)
First-stage statistic	164.05	61.72	30.22	164.05	61.72	30.22
<i>Panel B: Logged distances</i>						
2010 $\ln(\text{Distance to RER station})$	-0.212* (0.059)		-0.245* (0.059)	-0.132* (0.038)		-0.156* (0.040)
1968 $\ln(\text{Distance to non-RER stat/ramp})$		-0.436* (0.137)	-0.508* (0.128)		-0.310* (0.092)	-0.356* (0.087)

(continues)

TABLE 3 (Continued)

Dependent variable:	1968–2010 $\Delta \ln(\text{Employment density})$		1968–2010 $\Delta \ln(\text{Population density})$	
	[1]	[2]	[3]	[4]
1968 $\ln(\text{Employment density})$	-0.553* (0.077)	-0.609* (0.083)	-0.583* (0.077)	-0.014 (0.057)
1968 $\ln(\text{Population density})$	0.626* (0.090)	0.630* (0.078)	0.443* (0.090)	0.027 (0.051)
First-stage statistic	154.04	41.71	37.92	154.04
Distance to the nearest 1968 center	Y	Y	Y	Y
Geography	Y	Y	Y	Y
History	Y	Y	Y	Y
1968 Socioeconomy	Y	Y	Y	Y
Instrument:				
Distance to the nearest 1870 railroad line	✓	✓	✓	✓
Distance to the nearest Roman road	✓	✓	✓	✓

Panel B: Logged distances

Notes: One thousand three hundred observations for each regression. Geography variables are land area, altitude, index of terrain ruggedness, and elevation range. History variables are the population level in 1962 and dummy variables for municipalities (1) that were Roman settlements (based on DARMC maps), (2) that were major towns between the 10th and the 15th centuries (based on DARMC maps), (3) that were major towns between the 16th and the 19th centuries (based on Bairoch, 1988), (4) with a monastery built between the 12th and 16th centuries (based on DARMC maps), and (5) that hosted important fairs between the 10th and the 16th centuries (based on DARMC maps). Socioeconomic variables are the 1968 unemployment rate, the 1968 shares of employment in Manufacturing, in Construction, and in Services, the 1968 share of executives and professionals, and the 1968 share of population with university degree. Robust standard errors are in parentheses. \*, \*\*, and \*\*\* indicates significant at 1, 5, and 10 percent level, respectively.

growth, we start by simply including the 2010 distance to an RER station in column 1; then, we only include the 1968 distance to a non-RER transportation mode in column 2; and, we eventually include both transportation variables in column 3 (all specifications additionally include the other control variables). The corresponding results for population growth are reported in columns 4, 5, and 6, respectively. We find negative and significant effects for RER and non-RER transportation, revealing that employment and population growth increase the closer a municipality is to a railroad station (RER and non-RER) or a highway ramp.

More precisely, the specifications including distances to both RER and non-RER infrastructures (columns 3 and 6) reveal that each additional kilometer closer to the nearest RER station increases employment and population growth by about 2 percent and 1 percent, respectively. The corresponding elasticities are of 2 percent and 1.5 percent for a 10 percent increase in distance to the closest RER station. Yet, the effects are higher for the 1968 non-RER transportation, which is found to increase employment and population growth by 9 percent and 6 percent, respectively; the corresponding elasticities being of 5 percent and 4 percent, respectively.<sup>7</sup> Finally, it is important to note that the coefficients for both distances are not statistically different when they are individually, as opposed to jointly, estimated (columns 1 and 2 vs. 3 for employment, columns 4 and 5 vs. 6 for population). We take advantage of this feature in our last empirical analysis.

A closer look at the estimates for initial employment and population densities also provides interesting insights to understand the dynamic of employment growth (columns 1 to 3). The positive and significant coefficients for the 1968 log of population density reveal that employment growth is higher in municipalities that were initially more densely populated. On the other hand, the negatively significant coefficients for the 1968 log of employment density show that employment growth is lower in municipalities which initially had a larger employment density. Combined together, these two results tell us that employment follows population.

### 3.2 | Proximity matters!

Admittedly, the results discussed above show limited growth effects. This is not surprising since, as we have already noticed, these effects are *average* effects estimated for the 1,300 municipalities that make up the Paris metropolitan area. In this subsection, we explore whether these effects are uniform or rather heterogeneous across space.

We start by grouping municipalities according to their proximity to an RER station. Since the average commuting distance in metropolitan Paris is around 11–13 km (Aguilera & Mignot, 2004; Aguilera, 2005) and walking and biking are key to rail use, we focus on the closest municipalities and consider five different concentric rings each 1 km larger than the previous, starting from a minimum distance of 1 km from the nearest RER station to as far as 5 km. Following Equation (2), we thus estimate the effect on employment and population growth of the distance to an RER station for five different subset of municipalities: 188 municipalities located less than 1 km from an RER station; 321 located less than 2 km away; 421 located less than 3 km away; 476 located less than 4 km away; and 530 located less than 5 km away. The corresponding results are displayed in Table 4, where density gradients are reported in Panel A and elasticities in Panel B. The estimated effects on employment (respectively population) are reported in columns 1 to 5 (respectively 6 to 10), each column corresponding to a different subset of municipalities (from less than 1 km away in columns 1 and 6 to less than 5 km away in columns 5 and 10). These results confirm that the effect decreases with distance from an RER station, and clearly reveal that it is particularly stronger in municipalities which are closer to an RER station: from 8 percent in the 1-km ring to 3 percent in the 5-km ring for employment growth, and from 6 percent to 2 percent for population growth. By contrast, the corresponding results over all municipalities were of about 2 percent for employment and 1 percent for population (columns 3 and 6 of Table 3, Panel A).

<sup>7</sup> We also conducted the estimations with the 2010 (log of) distance to the nearest non-RER transportation in specifications corresponding to those in columns 2, 3, 5, and 6. The estimated coefficients for the (log) 2010 non-RER distance are statistically indistinguishable from those for the (log) 1968 distance reported in Table 3.

**TABLE 4** The effect of RER on municipality growth, TSLS: The closest municipalities

Dependent variable:	1968–2010 $\Delta \ln(\text{Employment density})$					1968–2010 $\Delta \ln(\text{Population density})$				
	dist to RER					dist to RER				
	$\leq 1$ km [1]	$\leq 2$ km [2]	$\leq 3$ km [3]	$\leq 4$ km [4]	$\leq 5$ km [5]	$\leq 1$ km [6]	$\leq 2$ km [7]	$\leq 3$ km [8]	$\leq 4$ km [9]	$\leq 5$ km [10]
2010 Distance to RER	-0.081* (0.023)	-0.063* (0.015)	-0.056* (0.015)	-0.042* (0.014)	-0.035** (0.014)	-0.062* (0.017)	-0.040* (0.012)	-0.031* (0.012)	-0.027* (0.010)	-0.024** (0.009)
First-stage statistic	20.12	38.25	43.61	54.05	59.20	20.12	38.25	43.61	54.05	59.20
2010 $\ln(\text{Dist to RER})$	-0.642* (0.206)	-0.754* (0.196)	-0.799* (0.194)	-0.643* (0.174)	-0.606* (0.167)	-0.522* (0.157)	-0.475* (0.144)	-0.473* (0.137)	-0.413* (0.112)	-0.394* (0.108)
First-stage statistic	21.56	26.42	31.82	44.68	50.33	21.56	26.42	31.82	44.68	50.33
1968 Dist to non-RER	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
1968 $\ln(\text{Densities})$	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Dist to 1968 center	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Geography	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
History	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
1968 Socioeconomy	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Instrument:										
Dist to 1870 railroad	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Dist to Roman road	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations:	188	321	421	476	530	188	321	421	476	530

Notes: Geography variables are land area, altitude, index of terrain ruggedness, and elevation range. History variables are the population level in 1962 and dummy variables for municipalities (1) that were Roman settlements (based on DARMC maps), (2) that were major towns between the 10th and the 15th centuries (based on DARMC maps), (3) that were major towns between the 16th and the 19th centuries (based on Bairoch, 1988), (4) with a monastery built between the 12th and 16th centuries (based on DARMC maps), and (5) that hosted important fairs between the 10th and the 16th centuries (based on DARMC maps). Socioeconomic variables are the 1968 unemployment rate, the 1968 shares of employment in manufacturing, in construction, and in services, the 1968 share of executives and professionals, and the 1968 share of population with university degree. Robust standard errors are in parentheses. \*, \*\*, and \*\*\* indicate significant at 1, 5, and 10 percent level, respectively.

We then group municipalities according to their proximity to employment (sub)centers. The idea is to compare municipalities in denser areas to those in peripheral areas because they might be different in their commuting patterns, types of employment and/or population, housing stock and tenure, among other characteristics. Table 5 reports results of estimating Equation (2) for employment (columns 1–4) and population (columns 5–8) when using direct distances (Panel A) and logged distances (Panel B). We split the 1,300 municipalities between the 221 located less than 5 km from an employment center (columns 1 and 2 and 5 and 6), and the remaining 1,079 located beyond 5 km (columns 3 and 4 and 7 and 8). Additionally, in some regressions, we drop municipalities that are part of (sub)centers (columns 2, 4, 6, and 8). In general, all results reinforce the previous idea that the effect of proximity to RER is higher in municipalities located in denser areas than in peripheral ones: 7 to 6 percent (columns 1 and 2 and 5 and 6) versus 2 to 1 percent (columns 3 and 4 and 7 and 8). Furthermore, in denser areas the effect is higher in noncentral municipalities: the estimated coefficients increase when we drop (sub)centers from the sample (columns 1 and 5 vs. columns 2 and 6).

In summary, although the results reported in Table 3 show limited *average* growth effects, we see that these effects are heterogeneous across space: the previous analysis clearly shows that growth effects are *locally higher* close to RER stations and employment (sub)centers. In other words, proximity matters!

### 3.3 | Employment and population types matter!

The Paris metropolitan area has undergone important demographic and socioeconomic changes between 1968 and 2010. Some of these changes are related to the type of employment and population. In terms of sector of activity, the share of total employment decreased from 24 to 6 percent in agriculture, from 14 to 11 percent in construction, and from 22 to 11 percent in manufacturing, while it increased from 40 to 72 percent in services. A similar trend can be observed when we look at occupations: the share of total jobs decreased from 12 to 4 percent for farmers, from 11 to 10 percent for craftsmen, and from 43 to 21 percent for factory workers, while it increased from 3 to 14 percent for executives and professionals, from 11 to 24 percent for intermediate occupations, and from 19 to 26 percent for employees. Regarding the education level as well, we observe a decrease in the share of uneducated individuals from 47 to 29 percent and in the share of low-educated individuals from 45 to 24 percent, while the share of population with high school degree and university degree increased from 5 to 18 percent and 3 to 29 percent, respectively. As a result, we now investigate the existence of heterogeneous effects according to the type of employment (sectors and occupations) and education level of population. Table 6 reports TSLS results for employment across sectors (agriculture, construction, manufacturing, services in columns 1 to 4), and occupations (farmers, craftsmen, executives and professionals, intermediates, employees, factory workers in columns 5 to 10), and for population across education levels (uneducated, low educated, high school degree, university degree in columns 11 to 14). As before, Panel A displays results obtained using direct distances and Panel B those using logged distances. A first look at the results reveals that the presence of an RER station has heterogeneous effects in this respect as well. First, while proximity to RER does not affect employment growth in agriculture and construction, getting closer to an RER station increases employment growth in manufacturing and in services especially. Second, only executives and professional workers, intermediate workers, employees, and factory workers benefit from proximity to an RER station. Finally, being closer to an RER station seems to systematically benefit population growth, with a slightly more marked effect for population with high school degree.

### 3.4 | Time matters!

We eventually investigate the potentially heterogeneous effect of RER proximity over time. The purpose is threefold. First, to estimate a version of Equation (2) in which our main explanatory variable, the distance to the nearest RER station, also uses values in the initial year. To consider this “more traditional” growth equation, we focus on the 1975–2010 period, which witnessed the advent and expansion of the RER network. This is a way of testing the robustness of the estimated effect of the RER. Second, we also investigate the temporal scope of the

**TABLE 5** The effect of RER on municipality growth, TSLS: Core vs. periphery

Dependent variable:	1968–2010 $\Delta \ln(\text{Employment density})$				1968–2010 $\Delta \ln(\text{Population density})$			
	dist to centers		dist to centers		dist to centers		dist to centers	
	$\leq 5$ km [1]	$> 5$ km [2]	$\leq 5$ km (no centers) [3]	$> 5$ km (no centers) [4]	$\leq 5$ km [5]	$\leq 5$ km (no centers) [6]	$> 5$ km [7]	$> 5$ km (no centers) [8]
Panel A: Direct distances								
2010 Distance to RER	-0.074* (0.028)	-0.125* (0.046)	-0.019* (0.005)	-0.019* (0.005)	-0.063* (0.018)	-0.091* (0.029)	-0.012* (0.003)	-0.012* (0.003)
First-stage statistic	4.57	3.21	31.56	30.40	4.57	3.21	31.56	30.40
Panel B: Logged distances								
2010 $\ln(\text{Dist to RER})$	-0.773** (0.324)	-0.935** (0.380)	-0.198* (0.059)	-0.203* (0.059)	-0.655* (0.222)	-0.654* (0.244)	-0.123* (0.041)	-0.127* (0.040)
First-stage statistic	2.95	4.06	33.45	31.94	2.95	4.06	33.45	31.94
1968 Dist to non-RER	Y	Y	Y	Y	Y	Y	Y	Y
1968 $\ln(\text{Densities})$	Y	Y	Y	Y	Y	Y	Y	Y
Dist to 1968 center	Y	Y	Y	Y	Y	Y	Y	Y
Geography	Y	Y	Y	Y	Y	Y	Y	Y
History	Y	Y	Y	Y	Y	Y	Y	Y
1968 Socioeconomy	Y	Y	Y	Y	Y	Y	Y	Y
Instrument:								
Dist to 1870 railroad	✓	✓	✓	✓	✓	✓	✓	✓
Dist to Roman road	✓	✓	✓	✓	✓	✓	✓	✓
Observations:	221	159	1,079	1,033	221	159	1,079	1,033

Notes: Geography variables are land area, altitude, index of terrain ruggedness, and elevation range. History variables are the population level in 1962 and dummy variables for municipalities (1) that were Roman settlements (based on DARMC maps), (2) that were major towns between the 10th and the 15th centuries (based on DARMC maps), (3) that were major towns between the 16th and the 19th centuries (based on Bairoch, 1988), (4) with a monastery built between the 12th and 16th centuries (based on DARMC maps), and (5) that hosted important fairs between the 10th and the 16th centuries (based on DARMC maps). Socioeconomic variables are the 1968 unemployment rate, the 1968 shares of employment in manufacturing, in construction, and in services, the 1968 share of executives and professionals, and the 1968 share of population with university degree. Robust standard errors are in parentheses. \*, \*\*, and \*\*\* indicates significant at 1, 5, and 10 percent level, respectively.

**TABLE 6** The effect of RER on municipality growth, TSLs: Sectors, occupations and education

Dependent variable:	1968–2010 $\Delta \ln(\text{Employment density})$										1968–2010 $\Delta \ln(\text{Population density})$			
	Sectors					Occupations					Education level			
	Agri	Con	Man	Ser		Farm	Craft	Exec	Int	Empl	Fact	None	Low	HSch
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]
2010 Distance to RER	0.003 (0.006)	-0.011 (0.007)	-0.018** (0.008)	-0.028* (0.006)	0.009*** (0.006)	-0.007 (0.006)	-0.025* (0.007)	-0.025* (0.007)	-0.022* (0.006)	-0.026* (0.007)	-0.013* (0.004)	-0.019* (0.003)	-0.028* (0.004)	-0.023* (0.006)
First-stage statistic	29.47	30.32	30.61	30.71	31.10	30.61	29.83	30.33	30.48	35.75	32.77	38.92	36.31	
2010 $\ln(\text{Dist to RER})$	0.019 (0.070)	-0.100 (0.079)	-0.174*** (0.093)	-0.307* (0.066)	0.109 (0.069)	-0.064 (0.071)	-0.257* (0.087)	-0.271* (0.078)	-0.231* (0.072)	-0.292* (0.075)	-0.134* (0.038)	-0.213* (0.035)	-0.298* (0.047)	-0.232* (0.061)
First-stage statistic	38.28	38.35	38.77	38.40	39.92	38.64	37.97	38.06	38.61	43.66	38.15	45.00	44.05	
1968 Dist to non-RER	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
1968 $\ln(\text{Densities})$	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Dist to 1968 center	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Geography	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
History	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
1968 Socioeconomy	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Instrument:														
Dist to 1870 railroad	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Dist to Roman road	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

Notes: One thousand three hundred observations for each regression. Geography variables are land area, altitude, index of terrain ruggedness, and elevation range. History variables are the population level in 1962 and dummy variables for municipalities (1) that were Roman settlements (based on DARMC maps), (2) that were major towns between the 10th and the 15th centuries (based on DARMC maps), (3) that were major towns between the 16th and the 19th centuries (based on Bairoch, 1988), (4) with a monastery built between the 12th and 16th centuries (based on DARMC maps), and (5) that hosted important fairs between the 10th and the 16th centuries (based on DARMC maps). Socioeconomic variables are the 1968 unemployment rate, the 1968 shares of employment in manufacturing, in construction, and in services, the 1968 share of executives and professionals, and the 1968 share of population with university degree. Robust standard errors are in parentheses. \*, \*\*, and \*\*\* indicates significant at 1, 5, and 10 percent level, respectively.

RER effects. As mentioned above, employment and population responses to transportation improvements might take years. Since the length of this delay is unclear, we explore it by regressing growth equations for the 1975–1990 and 1990–2010 subperiods. Finally, as shown in Table B.1, the different RER lines were built in separate periods: while in 1975 there was only one line, in 1990 there were three additional lines. As a result, this temporal analysis allows us to compare the effect of the whole network in 2010 with the effects of the RER lines existing in 1975 and 1990.<sup>8</sup>

To this aim, we regress the employment and population growth between *year t – 1* and *t* on the distance to the nearest RER station in *year t – 1* (or *t – 2*), conditional on employment and population densities in *t – 1*, and on distance to the nearest employment center, geography, history, and socioeconomic variables in *t – 1*:

$$\begin{aligned}
 (\text{year } t - 1 \text{ to year } t) \Delta \ln(\text{density}) = & \eta_0 + \eta_1 \times \text{year } t - 1 \text{ or } t - 2 f(\text{distance to RER station}) \\
 & + \eta_2 \times \text{year } t - 1 \ln(\text{densities}) \\
 & + \eta_3 \times f(\text{dist to the nearest 1968 empl. center}) \\
 & + \sum_i (\eta_{4,i} \times \text{geography}_i) + \sum_i (\eta_{5,i} \times \text{history}_i) \\
 & + \sum_i (\eta_{6,i} \times \text{year } t - 1 \text{ socioeconomy}_i). \tag{3}
 \end{aligned}$$

Compared to Equation (2), Equation (3) omits the distance to other transportation infrastructures since we observed (at the end of Section 2.2) that it does not significantly affect the coefficients of interest. This empirical strategy also allows us to overcome a problem with one of our instruments, the distance to the nearest Roman road, which is not relevant in some periods. As a result, Equation (3) is estimated by TSLS using the distance to the nearest 1870 railroad as the unique instrument.

Table 7 reports the results for employment (columns 1–4) and population (columns 5–8), alternatively using direct distances (Panel A) and logged distances (Panel B): columns 1 and 5 display estimates over the shortened 1975–2010 period, while the estimates in columns 2 and 6 are obtained on the first subperiod (1975–1990) and those in columns 3, 4, 7, and 8 correspond to the latest subperiod (1990–2010). Finally, while most specifications follow the traditional growth equation with the *year t – 1* RER distance (columns 1–3 and 5–7), in columns 4 and 8 we use the *year t – 2* RER distance (1975). Results for the whole 1975–2010 period (columns 1 and 5) are consistent with those obtained previously in Table 3: proximity to an RER station in 1975 increases employment and population growth. While RER effects when using direct distances are similar to their counterparts in Table 3, columns 3 and 6, the effects for logged distances are higher.

More interestingly, the analysis by subperiod shows different time responses to RER improvements for employment and population. For the case of employment, we do not find a significant RER effect on the 1975–1990 period (column 2). By contrast, the RER effect appears in the 1990–2010 period when we use the 1990 distance in column 3 (which includes both the 1975 and 1975–1990 RER networks) and it is clearer when we use the 1975 distance in column 4. Therefore, it seems that the RER effect emerges after a certain time lag. As for population, the RER effect turns out to be much more rapid, increasing population growth in the 1975–1990 period (column 6). However, at the same time, the RER effect tends to decrease over time: increasing population growth only by 0.3 percent (column 7) and 0.5 percent (column 8) in the 1990–2010 period.

To sum up, investigating the RER effects across different time periods confirms the robustness of our previous results based on the 1968–2010 period. Most importantly, this analysis reveals that these effects are heterogeneous across time periods and differ between firms and households: while there is a lagged response by firms which increases with time, the response in terms of residential location is more rapid, but decreases with time.

TABLE 7 The effect of RER on municipality growth, TSLS: Effects by periods

Dependent variable:	Δln(Employment density)		Δln(Population density)					
	1975-2010 [1]	1975-1990 [2]	1990-2010 [3]	1975-1990 [4]	1975-2010 [5]	1975-1990 [6]	1990-2010 [7]	[8]
Panel A: Direct distances								
1975 Dist to RER station	-0.027* (0.008)	-0.006 (0.007)	-0.022* (0.008)		-0.018* (0.004)	-0.014* (0.004)		-0.005*** (0.003)
1990 Dist to RER station			-0.012* (0.004)				-0.003*** (0.001)	
First-stage statistic	50.43	50.43	76.19	27.07	50.43	50.43	76.19	27.07
Panel B: Logged distances								
1975 Dist to RER station	-1.215* (0.458)	-0.242 (0.327)		-1.301** (0.622)	-0.857* (0.261)	-0.644* (0.220)		-0.344*** (0.182)
1990 Dist to RER station			-0.222** (0.087)				-0.050*** (0.029)	
First-stage statistic	17.84	17.84	43.09	6.91	17.84	17.84	43.09	6.91
1975 or 1990 ln(Densities)	Y	Y	Y	Y	Y	Y	Y	Y
Dist to 1968 center	Y	Y	Y	Y	Y	Y	Y	Y
Geography	Y	Y	Y	Y	Y	Y	Y	Y
History	Y	Y	Y	Y	Y	Y	Y	Y
1975 or 1990 Socioeconomy	Y	Y	Y	Y	Y	Y	Y	Y
Instrument:								
Dist to 1870 railroads	✓	✓	✓	✓	✓	✓	✓	✓

Notes: One thousand three hundred observations for each regression. Geography variables are land area, altitude, index of terrain ruggedness, and elevation range. History variables are population levels in 1962 and 1968 (all columns) and 1975 and 1982 (columns 3 and 6); employment levels in 1968 (all columns) and 1975 and 1982 (columns 3 and 6); and dummy variables for municipalities (1) that were Roman settlements (based on DARMC maps), (2) that were major towns between the 10th and the 15th centuries (based on DARMC maps), (3) that were major towns between the 16th and the 19th centuries (based on Bairoch, 1988), (4) with a monastery built between the 12th and 16th centuries (based on DARMC maps), and (5) that hosted important fairs between the 10th and the 16th centuries (based on DARMC maps). Socioeconomic variables are the unemployment rate, the shares of employment in manufacturing, in construction, and in services, the share of executives and professionals, and the share of population with university degree with their values in 1975 (columns 1-2 and 4-6) or in 1990 (columns 3 and 6). Robust standard errors are in parentheses. \*, \*\*, and \*\*\* indicates significant at 1, 5, and 10 percent level, respectively.

## 4 | CONCLUSIONS

In this paper, we investigated the effect of the RER expansion on employment and population growth in the municipalities of the Paris metropolitan area between 1968 and 2010, controlling for all other transportation modes. The main results indicate that the RER network together with the other transportation networks have a positive and significant effect on the location of employment and population: we find that each kilometer closer to an RER station increases employment and population growth by 2 percent and 1 percent, respectively. Analyzing the heterogeneity of these results across space, employment, and population types shows that these effects are higher (1) for municipalities, which are the closest to an RER station and the closest to an employment (sub)center; and (2) for services, for factory workers, and for population with high school and university degrees. Finally, looking at the heterogeneity over time reveals that there is no impact of the RER expansion on employment growth during the first part of the covered period, while the impact on population growth was sizable much earlier but declined over time.

This paper's contribution is nonnegligible as it provides much-needed evidence from an analysis conducted at the intrametropolitan level in one of the largest metropolitan areas in Europe. Furthermore, our results for the Paris metropolitan area complement those obtained by Mayer and Trévien (2015) using a different empirical strategy with a restricted sample of municipalities from the Paris metropolitan area. It is also important to note that some of our suburban and intrametropolitan results verify the theoretical predictions we discussed. First, we confirm that railroad and highway effects are heterogeneous in terms of distance to CBD. Second, we also provide evidence that the suburbanized population and employment are not evenly distributed across the suburbs: on the contrary, the population spreads out along the highways in the first stage of infrastructure development, while in this same stage, employment follows population.

A better understanding of the relationship between improvements in transport infrastructure, on the one hand, and city structure and city growth, on the other, is important, in general, for transport planners, urban planners, and policy makers and, in particular, it is crucial for making correct transport forecasts. Here, we have examined the impact of the initial stages in the development of the RER rail network on growth. Our results show that railroad investment has a major impact in these early years on population growth but not on employment growth. Further research is however required to determine whether this effect is weakened as the network becomes denser.

## ACKNOWLEDGMENTS

We are very grateful to Gilles Duranton, Laurent Gobillon, Miren Lafourcade, Daniel McMillen, Henry Overman, Jacques-François Thisse, and Jean-Claude Prager, as well as seminar and conference participants and two anonymous referees for their comments and suggestions. We also want to thank Corentin Trévien and Thierry Mayer for their very helpful GIS maps and data on the transportation networks. Financial support from the Société du Grand Paris, the Ministerio de Ciencia e Innovación (research projects ECO2010-20718 and ECO2014-52999-R [M.A. Garcia-López], and ECO2010-16934 and ECO2013-41310-R [E. Viladecans-Marsal]), Generalitat de Catalunya (research projects 2014SGR1326 [M.A. Garcia-López] and 2014SGR420 [E. Viladecans-Marsal]), and the "Xarxa de Referència d'R+D+I en Economia Aplicada" is gratefully acknowledged.

## REFERENCES

- Aguilera, A. (2005). Growth in commuting distances in French polycentric metropolitan areas: Paris, Lyon and Marseille. *Urban Studies*, 42(9), 1537–1547.
- Aguilera, A., & Mignot, D. (2004). Urban sprawl, polycentrism and commuting. A comparison of seven French urban areas. *Urban Public Economics Review*, 1, 93–114.
- Alonso, W. (1964). *Location and land use. Toward a general theory of land rent*. Cambridge, MA: Harvard University Press.
- Anas, A., & Moses, L. N. (1979). Mode choice, transport structure and urban land use. *Journal of Urban Economics*, 6(2), 228–246.
- Bairoch, P. (1988). *Cities and economic development: From the dawn of history to the present*. Chicago: University of Chicago Press.
- Banerjee, A., Dufo, E., & Qian, N. (2012). On the road: Access to transportation infrastructure and economic growth in China. Working Paper 17897, National Bureau of Economic Research.

- Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2), 775–805.
- Baum-Snow, N., & Kahn, M. E. (2000). The effects of new public projects to expand urban rail transit. *Journal of Public Economics*, 77(2), 241–263.
- Baum-Snow, N., & Kahn, M. E. (2005). Effects of urban rail transit expansions: Evidence from sixteen cities, 1970–2000. *Brookings-Wharton Papers on Urban Affairs*, 147–206.
- Bowes, D. R., & Ihlanfeldt, K. R. (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics*, 50(1), 1–25.
- Carlino, G. A., & Saiz, A. (2008). *City beautiful*. Working Paper 08–22, Federal Reserve Bank of Philadelphia.
- Chatman, D. G., & Noland, R. B. (2014). Transit service, physical agglomeration and productivity in us metropolitan areas. *Urban Studies*, 51(5), 917–937.
- Chen, Y., & Whalley, A. (2012). Green infrastructure: The effects of urban rail transit on air quality. *American Economic Journal: Economic Policy*, 4(1), 58–97.
- Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6), 2543–2594.
- Craig, S. G., Kohlhase, J. E., & Perdue, A. W. (2016). Empirical polycentricity: The complex relationship between employment centers. *Journal of Regional Science*, 56(1), 25–52.
- Craig, S., & Ng, P. T. (2001). Using quantile smoothing splines to identify employment subcenters in a multicentric urban area. *Journal of Urban Economics*, 49(1), 100–120.
- Diao, M., Zhu, Y., & Zhu, J. (2016). Intra-city access to inter-city transport nodes: The implications of high-speed-rail station locations for the urban development of Chinese cities. *Urban Studies*, Forthcoming.
- Donaldson, D. (2015). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, Forthcoming.
- Duranton, G., & Puga, D. (2015). Urban land use. In G. Duranton, J. V. Henderson & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 467–560). Amsterdam: North-Holland.
- Duranton, G., & Turner, M. A. (2011). The fundamental law of road congestion: Evidence from US cities. *American Economic Review*, 101(6), 2616–2652.
- Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. *The Review of Economic Studies*, 79(4), 1407–1440.
- Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies*, 81, 1046–1070.
- García-López, M.-A. (2010). Population suburbanization in Barcelona, 1991–2005: Is its spatial structure changing? *Journal of Housing Economics*, 19(2), 119–132.
- García-López, M.-A. (2012). Urban spatial structure, suburbanization and transportation in Barcelona. *Journal of Urban Economics*, 72(2–3), 176–190.
- García-López, M.-A., Holl, A., & Viladecans-Marsal, E. (2015). Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *Journal of Urban Economics*, 85(0), 52–67.
- García-López, M.-A., Pasidis, I., & Viladecans-Marsal, E. (2015). *Express delivery to the suburbs? Transportation effects in heterogeneous European cities*. Mimeo, IEB Working Paper 2015/30.
- Gibbons, S., & Machin, S. (2005). Valuing rail access using transport innovations. *Journal of Urban Economics*, 57(1), 148–169.
- Giuliano, G., Redfearn, C., Agarwal, A., Li, C., & Zhuang, D. (2007). Employment concentrations in Los Angeles, 1980–2000. *Environment and Planning A*, 39(12), 2935–2957.
- Giuliano, G. & Small, K. A. (1991). Subcenters in the Los Angeles region. *Regional Science and Urban Economics*, 21(2), 163–182.
- Glaeser, E. L., & Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature*, 47(4), 983–1028.
- Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. *Journal of Economic Geography*, 1(1), 27–50.
- Grube-Cavers, A., & Patterson, Z. (2015). Urban rapid rail transit and gentrification in Canadian urban centres: A survival analysis approach. *Urban Studies*, 52(1), 178–194.
- Hsu, W.-T., & Zhang, H. (2014). The fundamental law of highway congestion revisited: Evidence from national expressways in Japan. *Journal of Urban Economics*, 81(C), 65–76.
- Hurst, N. B., & West, S. E. (2014). Public transit and urban redevelopment: The effect of light rail transit on land use in Minneapolis, Minnesota. *Regional Science and Urban Economics*, 46(C), 57–72.

- Jiwattanakulpaisarn, P., Noland, R. B., Graham, D. J., & Polak, J. W. (2009). Highway infrastructure investment and county employment growth: A dynamic panel regression analysis. *Journal of Regional Science*, 49, 263–286.
- Kotavaara, O., Antikainen, H., & Rusanen, J. (2011). Population change and accessibility by road and rail networks: GIS and statistical approach to Finland 1970–2007. *Journal of Transport Geography*, 19(4), 926–935.
- Li, S., Yang, J., Qin, P., & Chonabayashi, S. (2016). Wheels of fortune: Subway expansion and property values in Beijing. *Journal of Regional Science*, 56(5), 792–813.
- Mayer, T., & Trévien, C. (2015). The impacts of urban public transportation: Evidence from the Paris metropolitan area. Working Paper G2015/03, INSEE.
- McDonald, J. F. (1987). The identification of urban employment subcenters. *Journal of Urban Economics*, 21(2), 242–258.
- McDonald, J. F., & Prather, P. J. (1994). Suburban employment centers: The case of Chicago. *Urban Studies*, 31(2), 201–218.
- McMillen, D. P. (2001). Nonparametric employment subcenter identification. *Journal of Urban Economics*, 50(3), 448–473.
- McMillen, D. P., & Redfearn, C. L. (2010). Estimation and hypothesis testing for nonparametric hedonic house price functions. *Journal of Regional Science*, 50(3), 712–733.
- Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, 90(4), 683–701.
- Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. *American Economic Review*, 57, 197–210.
- Moretti, E. (2004). Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review*, 94(3), 656–690.
- Muniz, I., Galindo, A., & García, M.-À. (2003). Cubic spline population density functions and satellite city delimitation: The case of Barcelona. *Urban Studies*, 40(7), 1303–1321.
- Muth, R. F. (1969). *Cities and housing: The spatial pattern of urban residential land use*. Chicago, IL: University of Chicago Press.
- Muñiz, I., Garcia-López, M.-À., & Galindo, A. (2008). The effect of employment sub-centres on population density in Barcelona. *Urban Studies*, 45(3), 627–649.
- Redfearn, C. L. (2007). The topography of metropolitan employment: Identifying centers of employment in a polycentric urban area. *Journal of Urban Economics*, 61(3), 519–541.
- Redfearn, C. L. (2009). How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets. *Regional Science and Urban Economics*, 39(3), 297–306.
- Rosenthal, S., & Strange, W. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2), 373–389.
- Ryan, S. (2005). The value of access to highways and light rail transit: Evidence for industrial and office firms. *Urban Studies*, 42(4), 751–764.
- Schuetz, J. (2015). Do rail transit stations encourage neighbourhood retail activity? *Urban Studies*, 52(14), 2699–2723.
- Sivitanidou, R. (1996). Do office-commercial firms value access to service employment centers? A hedonic value analysis within polycentric Los Angeles. *Journal of Urban Economics*, 40(2), 125–149.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Garcia-López M.-À, Hémet C, Viladecans-Marsal E. How does transportation shape intrametropolitan growth? An answer from the Regional Express Rail. *J Regional Sci.* 2017;00:1–23. <https://doi.org/10.1111/jors.10001>

## **D.8 Next train to the polycentric city: The effect of railroads on subcenter formation**



## Next train to the polycentric city: The effect of railroads on subcenter formation<sup>☆</sup>



Miquel-Àngel Garcia-López<sup>a,c,\*,1</sup>, Camille Hémet<sup>b,c,d,2</sup>, Elisabet Viladecans-Marsal<sup>c,d,e,3</sup>

<sup>a</sup> Universitat Autònoma de Barcelona, Spain

<sup>b</sup> Paris School of Economics, France

<sup>c</sup> Institut d'Economia de Barcelona, Spain

<sup>d</sup> Center for Economic Policy Research, United Kingdom

<sup>e</sup> Universitat de Barcelona, Spain

### ARTICLE INFO

#### JEL classification:

R11  
R12  
R14  
R4  
O2

#### Keywords:

Urban spatial structure  
Decentralization  
Subcenters  
Polycentric city  
Transportation

### ABSTRACT

Recent evidence reveals that transportation's improvements within metropolitan areas have a clear effect on population and job decentralization processes. Yet, very little has been said on how these improvements affect the spatial organization of the economic activity in the suburbs. This paper analyses the effects of transportation's changes on employment subcenters formation. Using data from metropolitan Paris between 1968 and 2010, we first show that the spatial pattern of job decentralization is reinforcing the polycentric nature of Paris: the number of subcenters grew from 21 in 1968 to 35 in 2010 and the employment growth was very intense within them. Second, our main contribution is to show that the new rail transit clearly affects the emergence of subcenters: not only does the presence of a rail station increase the probability of a suburban municipality of becoming (part of) a subcenter by 5 to 10%, but a 10% increase in municipality proximity to a suburban station is found to increase its chance to be part of a subcenter by 3 to 5%.

### 1. Introduction

Over the past fifty years, the Paris metropolitan area has been undergoing an unprecedented process of employment decentralization, whereby the employment share of the central city (Paris) fell from more than 45% in 1968 to less than a third today. In the meantime, national and regional governments have dedicated huge amount of money to improve public transportation in the area, with particular attention paid to the rail transit network. Suggestive evidence of this investment is provided by the Regional Express Rail (*Réseau Express Régional* in French, RER henceforth), a new and more efficient regional railway network that started operating in late 1960s and represented about

600 kilometers of rails in 2010. The main goal of this paper is thus to investigate the role played by the improvement of rail transit between 1968 and 2010 on the emergence of employment subcenters in Paris.

We organize our investigation in two parts. Departing from a previous work in which we found that Paris' job decentralization and, in particular, its intrametropolitan growth is shaped by the rail transit (see Garcia-López et al., *Forthcoming*), we first briefly describe metropolitan Paris and its rail transit, and analyze the spatial pattern of job decentralization: Is Paris decentralization diffuse or clustered around subcenters? McMillen's non-parametric approach (McMillen, 2001) allows us to identify employment subcenters in all six census years. The number of subcenters grew from 21 in 1968 to 35

<sup>☆</sup> We are very grateful to Gilles Duranton, Laurent Gobillon, Miren Lafourcade, Daniel P. McMillen, Henry Overman, Jacques-François Thisse, Matthew Turner, and Jean-Claude Prager, as well as seminar and conference participants and two anonymous referees for their comments and suggestions. We also want to thank Corentin Trévien and Thierry Mayer for their very helpful GIS maps and data on the transportation networks. Financial support from the Société du Grand Paris, the Ministerio de Ciencia e Innovación (research projects ECO2010-20718 and ECO2014-52999-R (M.A. Garcia-López), and ECO2013-41310-R and ECO2016-75941R (E. Viladecans-Marsal)), Generalitat de Catalunya (research projects 2014SGR1326 (M.A. Garcia-López) and 2014SGR420 (E. Viladecans-Marsal)), and the "Xarxa de Referència d'R+D+I en Economia Aplicada" is gratefully acknowledged.

\* Correspondence to: Department of Applied Economics, Universitat Autònoma de Barcelona, Edifici B, Facultat d'Economia i Empresa, 08193 Cerdanyola del Vallès, Spain.

E-mail addresses: [miquelangel.garcia@uab.cat](mailto:miquelangel.garcia@uab.cat) (M.-À. Garcia-López), [camille.hemet@psemail.eu](mailto:camille.hemet@psemail.eu) (C. Hémet), [eviladecans@ub.edu](mailto:eviladecans@ub.edu) (E. Viladecans-Marsal).

URLS: <http://gent.uab.cat/miquelangelgarcialopez> (M.-À. Garcia-López), <http://sites.google.com/site/camillehemet> (C. Hémet), <http://elisabetviladecans.wordpress.com> (E. Viladecans-Marsal).

<sup>1</sup> Department of Applied Economics, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain.

<sup>2</sup> Ecole Normale Supérieure (PSE), 48 Boulevard Jourdan, 75014 Paris, France.

<sup>3</sup> Department of Economics, Universitat de Barcelona, Avinguda Diagonal 690, 08034 Barcelona, Spain.

in 2010, some municipalities emerging as (part of) subcenters over the period while others were dropped out during the process. More importantly, our analysis reveals that employment growth in the subcenters during this period was very intense, both in absolute and relative terms. As a result, it seems clear that the spatial pattern of job decentralization in Paris is reinforcing the polycentric nature of its urban spatial structure.

Second, we investigate the role played by transportation on the emergence of those employment subcenters: Does rail transit cause subcenter formation? Our results show that the answer is ‘yes’: (1) the presence of a rail station increases the probability of a suburban municipality of being (part of a) subcenter by 5%, and (2) a 10% increase in municipality proximity to a suburban station causes about a 3% increase in its probability of being (part of) a subcenter. These results are robust to subcenter size and definition, and we only find that the effects are heterogeneous in terms of the type of rail: the suburban train and the Regional Express Rail (RER). While the effects for suburban train are similar to the average results mentioned above, the effect of the RER are much higher: the presence of a RER station increases the probability of being (part of) a subcenter by 10%, and the corresponding effect of getting 10% closer to a RER station amounts to 5%.

Our investigation contributes to the literature in three ways. As far as we know, we are the first to simultaneously study employment decentralization and subcenter formation in a very long time period. As Duranton and Puga (2015) point out, very little is known about the details of the spatial patterns of decentralized employment<sup>4</sup> and the frontier of knowledge was defined 15 years ago by Glaeser and Kahn (2001) and McMillen and Stefani (2003). Using data at the county level for 335 US cities, Glaeser and Kahn (2001) show that job decentralization between 1950 and 1990 was mainly diffuse. On the contrary, McMillen and Stefani (2003) use 1990 data at the Transportation Analysis Zone level to identify employment subcenters in 62 US cities and find that they are mainly polycentric<sup>5</sup>. In our paper, we analyze trends in job decentralization in a non-US city, Paris, and track the emergence of its employment subcenters from 1968 to 2010 using data at a fine spatial scale, the municipality. Our results reveal that the recent spatial pattern of job decentralization in Paris have reinforced its polycentric spatial configuration that was already apparent in 1968.

Second, this paper is the first to empirically study the role played by transportation on subcenter formation and thus the first to provide empirical evidence that supports theoretical models of urban spatial structure. As is well known, transportation plays a crucial role in the spatial distribution of residences and firms within cities. In the classical monocentric city model, transportation (accessibility) is the main factor that determines urban land use (Duranton and Puga, 2015). In non-monocentric models, the emergence of subcenters (and their number) depends on the interplay between agglomeration economies, transportation and population (Fujita and Ogawa, 1982; Helsley and Arthur, 1991; Henderson and Mitra, 1996; Henderson and Slade, 1996; Berliant et al., 2002; Lucas and Rossi-Hansberg, 2002; Anas and Kim, 1996; Berliant and Wang, 2008). From an empirical point of view, McMillen and Stefani (2003) are the only ones to explore the connection between transportation and subcenters. Because their work is restricted to 1990, they can only focus on the number of subcenters

and not on subcenter formation. Furthermore, since the number of subcenters is arguably determined simultaneously with transportation (and metropolitan population), this paper provides an interesting description of the data but not an estimate of causal effects as noted by Duranton and Puga (2015). In our paper, we study the causal effect of transportation on subcenter formation by using decennial census data from 1968 to 2010 to track transportation improvements and the emergence of subcenters in metropolitan Paris. We follow Duranton and Matthew (2012) and address endogeneity concerns relying on Instrumental Variables (IV) techniques with historical instruments built on the 19th century railroads (1870). Our results confirm this causality.

A final contribution of our research is further our understanding of the role of transportation infrastructure on shaping cities. At the intermetropolitan level, we know that transportation fosters urban growth (Duranton and Matthew, 2012), population suburbanization (Baum-Snow, 2007; Baum-Snow et al., 2015; García-López et al., 2015), employment decentralization (Baum-Snow, 2010, 2014), and modify zoning (García-López et al., 2015). At the intrametropolitan level, transportation affects employment and population densities (McMillen and McDonald, 1998; Redfearn, 2009a; García-López, 2010, 2012; Brooks and Lutz, 2016), local growth (García-López, 2012; Mayer and Trévién, 2015; García-López et al., Forthcoming), the housing market and the property values (Baum-Snow and Kahn, 2000; Bowes and Keith, 2001; McMillen and McDonald, 2004; Baum-Snow and Kahn, 2005; Gibbons and Machin, 2005; Ryan, 2005; Redfearn, 2009b; McMillen and Redfearn, 2010; Diao et al., 2016; Li et al., 2016), land use (Hurst and West, 2014; Schuetz, 2015), air pollution (Chen and Whalley, 2012) or cities’ gentrification (Grube-Cavers and Patterson, 2015). Our paper distinguishes itself from this large literature by showing that transportation also influences the spatial pattern of decentralized employment within cities by fostering the emergence of employment subcenters, a topic that remains largely unexplored, to the best of our knowledge.

The paper is organized as follows. Section 2 describes the main features of job location and railroad changes in the Paris metropolitan area over the past fifty years. Section 3 explores the pattern of this process, revealing that it has been clustered rather than diffused. Section 4 analyses the influence of railroad transit on the employment subcenters formation and finally Section 5, summarizes and concludes.

## 2. The Paris metropolitan area

France is organized into *régions*, administrative units with a governing body elected by direct vote (the Regional Council). Regions are divided into *départements*, also with a government (the General Council of the Department). Departments are made up of *communes* (municipalities), the lowest level of the administrative division governed by the City Council. Some big municipalities are subdivided into *arrondissements*. In terms of planning, the national government sets the legal framework, but it does not establish a national spatial plan. On the contrary, regions and municipalities are responsible of planning policies: the former prepare the general strategic plan (for example, infrastructure plans) and the latter develop local land use plans and approve building permits. While in the past, regional plans provided only non-binding guidelines and strategies, after November 2014 local land use plans need to adopt the logic of the regional plan (OECD, 2017).

This study focuses on the Paris metropolitan area, a French administrative region known as *Ile de France*. It is the densest and most populated metropolitan area in France, with 9.8 inhabitants per hectare in 2010 for a total of 11,786,234 inhabitants and 1,200,000 hectares. It is also the region with the highest employment density, with a total of 5,668,902 jobs in 2010 (21.6% of French employment). From a sector point of view, 0.2% of metropolitan jobs are in the agricultural sector, 5.2% in the construction sector, 8.4% in industry

<sup>4</sup> Most papers characterize the intrametropolitan location of employment by identifying subcenters and/or estimating density functions: while Giuliano and Small (1991); McDonald and Prather (1994) and McMillen (2001, 2004) show the existence of employment subcenters in some US cities, Glaeser and Kahn (2004) consider that job decentralization is mainly diffuse and directly related with cars. Others study the determinants of intrametropolitan growth (e.g. Bollinger and Ihlanfeldt (1997, 2003); García-López and Muñiz (2013)) and of growth in (within) subcenters (e.g. Giuliano and Small, 1999; Giuliano et al., 2012).

<sup>5</sup> Similar results are obtained by Arribas-Bel and Sanz-Gracia (2014) using local indicators of spatial association (LISA) to identify subcenters in 359 MSAs with census tract data for 1990, 2000 and 2010.

and the remaining 86.2% in the tertiary sector (trade, services, public administration, education, among others) (García-López et al., [Forthcoming](#)).

Metropolitan Paris is composed of its CBD, the city of Paris (with 20 *arrondissements*), and 1280 suburban municipalities. The average land size of the CBD is 526 ha (with a standard deviation of 435 ha). For suburban municipalities, the average size is 724 ha (466 ha SD) for the ones located less than 25 km from Paris (average distance to the CBD), and it is 1146 ha (953 ha SD) for those located beyond the 25 km. In our empirical strategy we take into account these differences in land size by using density variables (as dependent or control variables).

Relying on detailed population and employment data at the municipal level from six census waves (1968, 1975, 1982, 1990, 1999 and 2010), we are able to track the evolution of the urban spatial structure of the Paris metropolitan area over the past forty years. We use municipal data because census tract data are only available from 1990<sup>6</sup>.

We also use precise transportation data, provided by Mayer and Trévien (2015) and the Institut d'Aménagement et d'Urbanisme (IAU), to characterize the changes in the area's public transportation over the period.

### 2.1. Employment decentralization in Paris

García-López et al. ([Forthcoming](#)) show that metropolitan Paris is undergoing a process of employment decentralization in which its central city (Paris) loses jobs in favor of suburban locations. In particular, the number of jobs in central Paris declined by 7.1% between 1968 and 2010. Over the same period, while the metropolitan area as a whole grew by about one third, the share jobs located in central Paris dropped from 45.3 to 31.7%. As a result, this evolution of jobs' location reveals that metropolitan Paris is decentralizing both in absolute and relative terms.

Additional evidence of this decentralization process comes from estimating the traditional monocentric density function for each census year. Using municipal data, we regress the log of employment density on the distance to the center of the CBD (Paris). In order to take nonlinearities into account, our estimations are based on a non-parametric method known as Locally Weighted Regression (LWR), with a bandwidth of 0.5<sup>7</sup> (McMillen, 2001)<sup>8</sup>. The LWR density estimates drawn in [Fig. 1](#) are for the whole city (top graph) and for four city quadrants. These graphs clearly illustrate the decentralization process between 1968 (blue line) and 2010 (red line). Indeed, between these two dates, employment density in the most central municipalities (0 to 10 km from the CBD) decreased (top and North-East graphs), did not change (South-West graph) or slightly increased (North-West and South-East graphs), while it increased greatly in municipalities located between 10 and 60 km from the CBD (all five graphs), in line with a (relative and/or an absolute) decentralization process. We can also note a reduction in employment density for the most peripheral (mostly rural) municipalities (more than 60 km from the CBD). This, combined with the increase observed for the suburban municipalities indicate the emergence of new suburban subcenters and reinforcement of existing ones around kilometers 30 (all five graphs), 40 (North-East and South-West graphs), and 50 and 60 (North-West graph).

<sup>6</sup> With a total area of 1,200,000 hectares and 1300 observations, Paris is closer to Los Angeles with 1146 tracts and 910,000 ha (see [Giuliano and Small \(1991\)](#)) than to Chicago with 14,290 tracts and 920,000 ha (see [McMillen and McDonald \(1998\)](#)). As [McMillen \(2001\)](#) points out, subcenter identification (in [Section 3](#)) depends on the size of the unit of observation: large observations may produce fewer subcenter sites than more disaggregated data.

<sup>7</sup> A window size of 0.5 means that the nearest 50% of observations are weighted.

<sup>8</sup> Results are similar when using window sizes of 0.1, 0.3, 0.8 and 1.0 and are available upon request.

### 2.2. Rail transit in Paris

As documented with more details in [García-López et al. \(Forthcoming\)](#), the public transportation infrastructure of the Paris metropolitan area has dramatically improved since the 1960s and today is mostly based on a railroad network made of more than 1600 km of lines and including four network types, as illustrated in [Fig. 2](#). First, a suburban train (henceforth *train*) connecting Paris to the rest of the metropolitan area (suburbs as well as some of the most remote rural municipalities), that underwent substantial improvement over the 1960s.

Second, the Paris region is endowed with a regional express network (*Réseau Express Régional* in French, RER henceforth) which started operating in late 1960s. Like the train, the RER connects Paris to the suburbs, but for a shorter maximum distance of about 30 km. Most of the RER lines follow the train lines and were designed to improve the existing train network. An important distinction between the train and RER networks is that the latter has connections within Paris. This means the RER enables passengers to commute from one part of the Paris Metropolitan Area to another, going through Paris, but without having to switch to another train to cross the city. This represents a clear improvement to regional transit overall. As a whole, the RER network increased its number of lines from 1 to 5, its total length from 39 to 587 km, its number of stations from 22 to 243, and its number of municipalities with stations from 16 to 167 between 1975 and 2010.

In addition to these regional railroad networks, Paris is characterized by a very dense subway system (*métro* henceforth), which started in 1900, and is mainly connecting areas within Paris. Between 1968 and 2010, the *métro* network kept expanding, such that a few *métro* stations are now located beyond Paris, in the immediate outskirts of the CBD.

The city of Paris and its closest suburban area (the first ring of municipalities out of Paris) also enjoy a tramway network. This fourth network is much more recent, dating back from the beginning of the 1990s, and is still expanding.

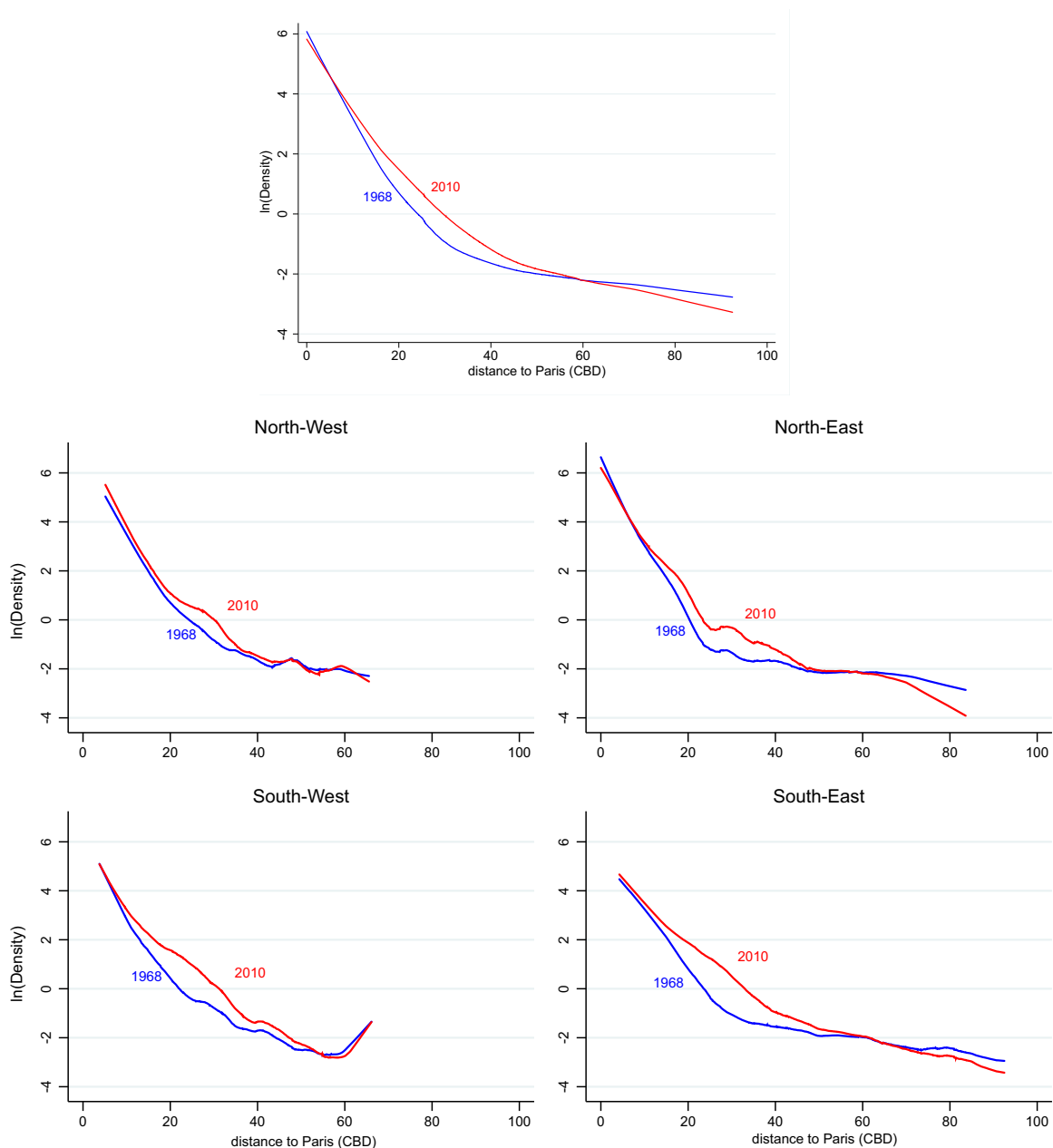
Finally, it is important to note that the origins of these rail transit networks can be traced back to the 19th century. In the empirical strategy of the following sections, we exploit this link to correct for the potential biases related to the endogenous location of rail stations and subcenter formation. In particular, we rely on IV techniques with two historical instruments as sources of exogenous variation: the distance to the nearest 1870 railroad line and a dummy variable indicating whether a given municipality was crossed by a railroad line in 1870. In [Appendix A](#) we extensively document and discuss the validity of these 1870 rail variables as instruments for the location of modern railroad stations.

## 3. Is Paris decentralization diffuse or clustered around subcenters?

After establishing that the Paris metropolitan area went through a job decentralization process between 1968 and 2010, we now want to characterize the spatial pattern of this process: Does decentralization follow a *polycentric* spatial pattern, reinforcing existing secondary centers (subcenters) and/or fostering the emergence of new ones? Or does it rather reflect a *dispersed* spatial pattern, in which suburban land is occupied by low-density settlements? To answer these questions, we first identify subcenters for each census year between 1968 and 2010 before analyzing the evolution of employment inside these subcenters vs. non-central locations between 1968 and 2010.

### 3.1. Identifying and characterizing subcenters

The empirical literature has proposed alternative procedures to identify subcenters. The most used are those based on density and



**Fig. 1.** Employment decentralization in metropolitan Paris, 1968–2010  
 Note: Density estimates based on LWR with a window size of 0.5.

employment thresholds (Giuliano and Small, 1991; Giuliano et al., 2007; Muñiz et al., 2008) and on employment and/or population density peaks (McDonald, 1987; McDonald and Paul, 1994; Craig and Pin, 2001; Muñiz et al., 2003; Redfearn, 2007; García-López, 2010; Craig et al., 2016). Despite using different methods<sup>9</sup>, they all agree that an employment subcenter is a place with a significantly larger employment density than nearby locations that has a significant effect on the overall spatial distribution of jobs.

We identify employment subcenters using the method first developed by McDonald and Prather (1994) and improved by

McMillen (2001) because it is consistent with theory and, in particular, based on (deviations from) the monocentric model: The main idea is to estimate densities following a monocentric spatial pattern; the predicted densities obtained are then subtracted from the corresponding real densities; and the positive and statistically significant residuals are finally selected.

While McDonald and Prather (1994) estimate by Ordinary Least Squares (OLS) a two dimensional density function (log of employment density on the distance to CBD), as in Fig. 1, McMillen (2001) suggests estimating a three-dimensional density function (log of employment density on north-south and east-west distances to CBD) with a Locally Weighted Regression (LWR). Both improvements allow to take into account geographical differences, which, in terms of the spatial pattern of densities, can occur in any direction from the CBD (e.g. steeper density gradients on the north side than on the south side of the city). They additionally allow to define any type of monocentric spatial pattern: concave, convex or linear.

<sup>9</sup> Although in general they identify different sets of subcenters in the same city (for instance, see Redfearn (2007) in Los Angeles when using his own methodology than when relying on Giuliano and Small (1991) and McMillen (2001)'s procedures), there are examples in which different methods lead to similar sets of subcenters (see for instance Craig et al. (2016) in Houston when using McMillen (2001) and Redfearn (2007)'s strategies).

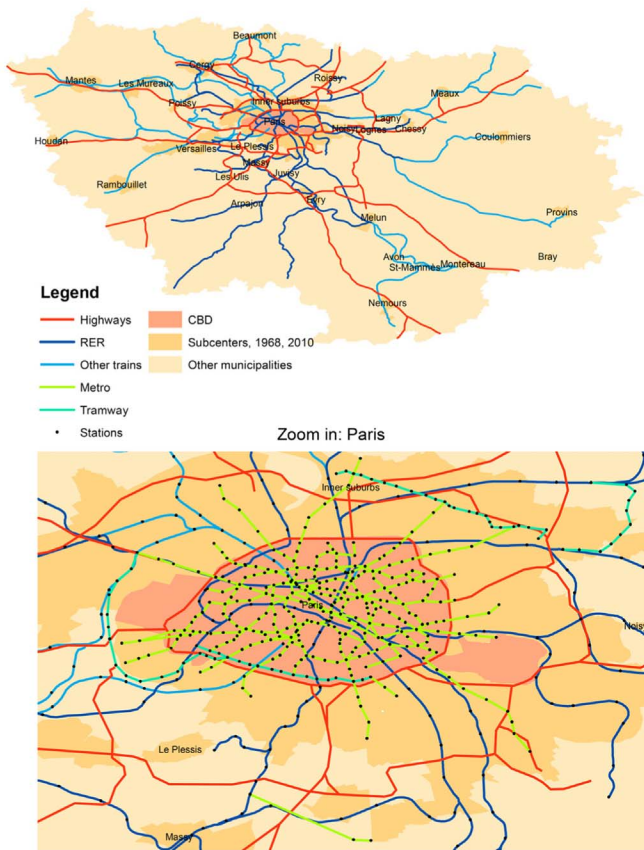


Fig. 2. Transportation infrastructures in metropolitan Paris, 2010.

We therefore estimate the following employment density function:

$$\ln(\text{Employment density}) = \gamma_0 + \gamma_1 \times \text{north-south distance to CBD} + \gamma_2 \times \text{east-west distance to CBD} \quad (1)$$

where density is measured as jobs per hectare, and distances are in kilometers. The CBD is defined as the 20 *arrondissements* that make up the city of Paris. Distance to CBD is the distance to the centroid of the 4<sup>th</sup> *arrondissement* (corresponding to the town-hall of Paris).

Since our estimates are based on LWR, we need to define a bandwidth. As [McMillen \(2001\)](#) points out, this is a critical choice because we need a monocentric benchmark. We experiment with alternative window sizes ranging from 1% to 9% and from 10% to 90% (see [Table B.1](#) in [Appendix B](#)). After visual inspection, we find that the first monocentric spatial structure appears when the nearest 50% observations are included in each local regression. Interestingly, this is the value used by [McMillen \(2001\)](#) for some US cities. We also experimented with a selection rule based on the Akaike information criterion. However, the selected window size (7%) was clearly related to a polycentric spatial structure (see and compare [Tables B.1](#) and [B.2](#) in [Appendix B](#)).

Second, for each site we compute the residual as the difference between the log of real employment density and the estimated log of employment density. We then select those that are significantly positive, according to their own standard errors that can vary over space ([McMillen, 2001](#)). We use two critical thresholds, 1.96 and 1.64, that are associated with a 5% and a 10% significance level, respectively.

Finally, we group the selected sites in subcenters when they are contiguous. We use a “queen” criterion for contiguity: two sites (municipalities) are contiguous if they share at least one point in their boundaries. See [McMillen \(2001, 2003\)](#) and [García-López \(2010\)](#) for further details on this procedure.

This methodology enables us to identify subcenters for each census

year between 1968 and 2010, which are described in [Table 1](#). For each year, we report two figures, corresponding to subcenters identified using positive residuals significant at the 5 and at the 10% level, respectively. From Panel A, we can see that the number of subcenters identified at the 5% level (respectively, at the 10% level) increased from 20 to 26 (respectively from 21 to 35) between 1968 and 2010. Panel B reveals that these subcenters hosted 1,756,000 jobs in 2010 (respectively 1,979,000), corresponding to an increase of about 600,000 jobs since 1968 (respectively 560,000). We can also observe that subcenters are heterogeneous in terms of size, with an increasing number of large subcenters (more than 20,000 jobs), in line with the decentralization process: in 1968, between 15 and 20% of subcenters hosted more than 20,000 jobs, contrasting with a range of 42 to 50% in 2010. Panel C shows the share of metropolitan employment in the CBD and the subcenters. While the CBD share decreased from 45% in 1968 to 32% in 2010, subcenters concentrated about a third of all city jobs during the studied period. Finally, we want to emphasize that the number of subcenters and the number of jobs in the subcenters do not differ much whether the subcenters are identified using positive residuals at the 5% level or at the 10% level. We will henceforth use the subcenters identified at the 10% level in our analysis.

[Table 2](#) further describes the identified subcenters (at the 10% level) regarding the number of municipalities they encompass: All of them in Panel A, with more than 10,000 jobs in Panel B, and with more than 20,000 jobs in Panel C. We can first notice that the number of municipalities that form a subcenter of their own or that are part of a subcenter is quite stable, from 88 (73, 67) in 1968 to 89 (76, 65) in 2010. We can also observe a certain stability in the composition of subcenters: among all the municipalities belonging to a subcenter, 57 (42, 38) are constantly identified as a subcenter (or as part of a subcenter) over time. For the case of Los Angeles, [Redfearn \(2009a\)](#) also finds evidence of the persistence of employment subcenters during the 20th century. The remaining municipalities may have emerged as part of a subcenter at some point, or, alternatively, stopped being considered as such. The last two lines of each Panel also report subcenters’ surface and show that 40 to 50% of land was associated with constantly identified subcenters in 2010. Furthermore, total subcenters’ land size increased by 20% between 1968 and 2010, despite their constant number. This indicates that the *new* subcenters are spatially larger.

### 3.2. The spatial pattern of decentralization

In order to determine the spatial pattern of employment decentralization, we now compare the evolution of employment inside and outside employment centers over the period of interest.

[Table 3](#) displays the number and the percentage of jobs located in each type of area for all census years (columns 1 to 6), and the corresponding variation between 1968 and 2010 (column 7). In Panel A, the figures correspond to the 20 *arrondissements* of Paris, the CBD. Panel B shows figures for different classifications of subcenters. The *All subcenters* category refers to subcenters identified at a given date, some of them appearing or disappearing as subcenters between two waves. In other words, the geography is not constant as the municipalities included in this category differ between two points in time (as shown in [Table 2](#) Panel A). By contrast, the *Always subcenters* and the *Subcenters in 2010 (not all years)* categories consider constant geography: the former by including the 57 municipalities that are identified as a subcenter (or part of one) in all six years, the latter by including the 32 municipalities identified as (part of) a subcenter in 2010, but not in all previous years. In Panel C, we assess the importance of employment centers by jointly considering the CBD and the *All subcenters* categories. Finally, figures for non-central municipalities are in Panel D. Similar to Panel B, geography is not constant in the *All non-central municipalities* category, and it is constant in the *Always non-central* and the *Non-central in 2010*

**Table 1**  
Employment subcenters in metropolitan Paris, 1968–2010.

	1968		1975		1982		1990		1999		2010	
	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
Resid. significant at:												
<b>Panel A: Number of subcenters</b>												
All subcenters	20	21	26	27	30	32	29	34	31	34	26	35
≥ 10,000 jobs	7	8	13	14	18	16	19	21	19	19	19	23
≥ 20,000 jobs	3	4	8	7	9	12	13	13	13	12	13	15
<b>Panel B: Jobs ('000) in subcenters</b>												
All subcenters	1152	1419	1319	1667	1237	1665	1373	1788	1441	1782	1756	1979
≥ 10,000 jobs	1088	1350	1254	1601	1169	1582	1326	1720	1369	1693	1714	1909
≥ 20,000 jobs	1022	1290	1184	1501	1040	1513	1252	1618	1281	1602	1624	1788
<b>Panel C: Metropolitan shares in the CBD and subcenters</b>												
CBD (Paris)	45.3%	45.3%	41.0%	41.0%	38.4%	38.4%	35.8%	35.8%	31.8%	31.8%	31.7%	31.7%
All subcenters	26.9%	33.2%	28.2%	35.7%	26.3%	35.4%	27.0%	35.2%	28.6%	35.3%	31.0%	34.9%
≥ 10,000 jobs	25.4%	31.6%	26.8%	34.2%	24.8%	33.6%	26.1%	33.9%	27.1%	33.6%	30.2%	33.7%
≥ 20,000 jobs	23.9%	30.2%	25.3%	32.1%	22.1%	32.2%	24.7%	31.9%	25.4%	31.8%	28.6%	31.5%

Note: LWR estimates use a window size of 0.5 (i.e., the nearest of the 50% observations).

**Table 2**  
Municipalities and land in employment subcenters in metropolitan Paris, 1968–2010.

	1968	1975	1982	1990	1999	2010
<b>Panel A: All subcenters</b>						
<i>Number of municipalities</i>						
All municipalities	88	97	93	95	95	89
Emerging as (part of) subcenters	–	18	7	11	12	5
Disappearing as (part of) subcenters	–	9	11	9	12	11
Always in subcenters	57	57	57	57	57	57
<i>Hectares of land</i>						
All municipalities	50,048	58,122	59,159	61,947	67,055	60,245
Always in subcenters	31,140	31,140	31,140	31,140	31,140	31,140
<b>Panel B: Subcenters ≥ 10,000 jobs</b>						
<i>Number of municipalities</i>						
All municipalities	73	81	75	80	78	76
Always in subcenters	42	42	42	42	42	42
<i>Hectares of land</i>						
All municipalities	41,038	49,086	48,998	54,296	57,582	51,514
Always in subcenters	21,814	21,814	21,814	21,814	21,814	21,814
<b>Panel C: Subcenters ≥ 20,000 jobs</b>						
<i>Number of municipalities</i>						
All municipalities	67	73	69	68	69	65
Always in subcenters	38	38	38	38	38	38
<i>Hectares of land</i>						
All municipalities	35,047	40,834	43,401	47,102	45,492	42,458
Always in subcenters	17,613	17,613	17,613	17,613	17,613	17,613

Note: Employment subcenters identified using McMillen (2001)'s method with a LWR window size of 50%, and for positive residuals significant at the 10% level.

(not all years) categories.

The figures reported in Table 3 confirm the decentralization process commented in Section 2: the CBD lost 138,000 jobs (–7%) between 1968 and 2010 and reduced its metropolitan share of jobs from 45% to 31% (Panel A).

The decentralization process and the overall employment growth experienced by the city benefited subcenters (Panel B). Despite always

concentrating around a third of all employment in the Paris metropolitan area, their jobs grew from 1.5 to 2 millions (39%) in the studied period. If we consider constant geographies, we can see that employment growth took place in the *Always subcenters* category, with 344,000 new jobs (33%) and, in particular, in those municipalities that ended up as subcenters in 2010 (the *Subcenters in 2010 (not all years)* category), with 500,000 new jobs (483%).

Similarly, both processes also benefited non-central municipalities (Panel D), which grew in 900,000 jobs (91%). While most growth took place in the *Always non-central* category (629,000 jobs, 83%), municipalities that ended up as non-central in 2010 only grew in 55,000 jobs (12%).

Finally, the importance of the polycentric nature of Paris is clear when we jointly consider all employment centers (Panel C) and compare with non-central municipalities (Panel D). In this sense, despite the reduction in their metropolitan share between 1968 and 2010, employment centers still concentrated 67% of jobs in 2010. On the contrary, non-central municipalities represented around one fifth of all jobs in the metropolitan area in 1968, but up to one third in 2010. In other words, while 3.8 millions of jobs were concentrated in 90 municipalities (89 subcenters + Paris), 1.8 millions of jobs were located in 1191 municipalities in 2010.

In summary, the increasing number of employment subcenters, their high employment growth rates and the fact that they concentrate a third of metropolitan jobs allow us to conclude that job decentralization in Paris is clustered around subcenters. We have to admit though that a small part of job decentralization is diffuse.

#### 4. Does rail transit cause subcenter formation?

After analyzing the job decentralization process in the Paris metropolitan area, we now turn to the most important part of this paper, where we contribute to the literature by establishing that rail transit causes subcenter formation. To answer to this key question, we proceed in two steps. We first investigate whether the existence of a rail station in a suburban municipality increases the probability that this municipality becomes (part of) a subcenter. Then, we examine whether proximity to rail stations also increases the likelihood of becoming (part of) a subcenter, even when the station is not built on the municipal ground.

In both steps, our empirical strategy consists in regressing the probability that a municipality becomes a subcenter on a *rail station variable*. In Section 4.1, where we explore the role of the existence of a rail station, this variable indicates whether there is a station within the administrative boundaries of the municipality or the number of stations

**Table 3**  
The spatial pattern of decentralized employment in metropolitan Paris, 1968–2010.

	1968 [1]	1975 [2]	1982 [3]	1990 [4]	1999 [5]	2010 [6]	1968–2010 [7]
<b>Panel A: CBD</b>							
Paris (20 <i>arrondissements</i> )	1936	1918	1808	1815	1601	1798	–138 (–7.1%)
Metropolitan share	45.3%	41.0%	38.4%	35.8%	31.8%	31.7%	
<b>Panel B: Subcenters</b>							
All subcenters	1419	1667	1665	1788	1782	1979	560 (39.5%)
Metropolitan share	33.2%	35.7%	35.4%	35.2%	35.3%	34.9%	
Always subcenters	1028	1136	1112	1168	1183	1372	344 (33.5%)
Metropolitan share	24.0%	24.3%	23.6%	23.0%	23.5%	24.2%	
Subcenters in 2010 (not all years)	104	221	316	438	506	606	502 (482.7%)
Metropolitan share	2.4%	4.7%	6.7%	8.6%	10.0%	10.7%	
<b>Panel C: All employment centers (A+B)</b>							
Paris + All subcenters	3355	3585	3473	3603	3383	3777	422 (12.6%)
Metropolitan share	78.4%	76.7%	73.8%	71.0%	67.1%	66.6%	
<b>Panel D: Non-central municipalities</b>							
All non-central municipalities	992	1090	1232	1472	1659	1893	901 (90.8%)
Metropolitan share	21.6%	23.3%	26.2%	29.1%	32.9%	33.4%	
Always non-central	754	880	961	1136	1260	1383	629 (83.4%)
Metropolitan share	17.6%	18.8%	20.4%	22.4%	25.0%	24.4%	
Non-central in 2010 (not all years)	455	518	508	519	493	510	55 (12.1%)
Metropolitan share	10.6%	11.1%	10.8%	10.2%	9.8%	9.0%	

Note: Employment values are thousands of jobs. Growth rates are in parentheses. Employment subcenters identified using [McMillen \(2001\)](#)'s method with a LWR window size of 50%, and for positive residuals significant at a 10% level.

and lines in the municipality. Alternatively, in [Section 4.2](#), this variable measures the distance between a municipality and the closest station.

We control for characteristics related to Paris urban spatial structure of metropolitan such the distance to the CBD, and employment and population densities. Also for municipal geography with altitude, index of terrain ruggedness, and elevation range variables. History variables are the population levels between 1962 and year  $t-1$  and dummy variables for municipalities (1) that were Roman settlements (based on DARMC<sup>10</sup> maps), (2) that were major towns between the 10th and the 15th centuries (based on DARMC maps), and (3) between the 16th and the 19th centuries (based on [Bairoch \(1988\)](#)), and (4) with a monastery built between the 12th and 16th centuries (based on DARMC maps). Socioeconomic variables are the unemployment rate, the shares of employment in manufacturing, in construction, and in services, the share of executives and professionals, and the share of population with university degree.

The general equation, that we estimate using probits, can thus be expressed as follows:

$$\begin{aligned}
 \text{Prob}(\text{subcenter}_{i,t}) = & \delta_0 + \delta_1 \times \text{rail station variable}_i \\
 & + \delta_2 \times \ln(\text{densities}_{i,t}) + \delta_3 \times \ln(\text{distance to CBD})_i \\
 & + \sum_i (\delta_{4,i} \times \text{geography}_i) + \sum_i (\delta_{5,i} \times \text{history}_i) \\
 & + \sum_i (\delta_{6,i} \times \text{socioeconomy}_{i,t}) \tag{2}
 \end{aligned}$$

In order to correct for the potential biases related to the endogenous location of rail stations, we use a dummy variable indicating whether a given municipality was crossed by a rail (a train line) in 1870 ([Section 4.1](#)) and the distance to the nearest 1870 railroad line ([Section 4.2](#)) and as instruments, as documented and discussed in [Appendix A](#). However, the use of these historical instruments comes with a caveat: since they are time invariant, we can not estimate using panel data techniques. As a result, we pool all observations together, irrespective

of the year, and include year fixed-effects in our regressions. Since only suburban trains operated in 1968 and RER started operating between late 1960 s and mid 1970 s, we show results for the 1968–2010 and the 1975–2010 periods. As a robustness, we also run some cross-sectional regressions.

#### 4.1. Do rail stations lead to subcenters?

In order to establish whether the existence of a rail station in a suburban municipality increases the probability that this municipality becomes (part of) a subcenter, we estimate [Eq. \(2\)](#) using the 1280 suburban municipalities (excluding the 20 *arrondissements* of Paris).

The marginal effects of the corresponding (second-stage) results are displayed in [Table 4](#). Panel A shows results when estimating by Ordinary Least Squares (OLS) and Panel B their Two-Stage Least Squares (TSLS) counterparts. Columns 1 to 7 presents results estimated using all suburban municipalities. In columns 1 and 2, the station variable represents the number of lines times station, which counts the total number of lines having a stop in a municipality (it can be seen as a weighted count of the number of stations). The station variable then simply counts the number of stations in columns 3 and 4, and indicates the existence of a station in columns 5 to 7.<sup>11</sup> We will restrict our comments on the specifications that includes all control variables (densities, distance to CBD, geography, history and socioeconomy), after noting that the marginal effects are significantly reduced in the conditional regressions<sup>12</sup>.

<sup>11</sup> Therefore, if a municipality has two stations, with  $n_1$  lines stopping in one station and  $n_2$  lines in the other one, the “number of lines-stations” variable takes a value of  $n_1 + n_2$ , the “number of stations” variable takes a value of 2, and the dummy variable is equal to 1.

<sup>12</sup> The importance of these control variables can be noticed in the evolution of the pseudo  $R^2$ : without controls, its value is below 20%; with them, it is close to 90%. The variables driving the big jump are employment density and population density. They increase pseudo  $R^2$  to 70%, something expected because the identification of subcenters use a job density criterion. Adding the distance to CBD, the explanatory power increases to 85%. Finally, historic, geographic and socioeconomic variables only add around 3 percent points. However, remember that these variables are included because our historical instruments need to fulfill the exclusion restriction (see [Appendix A](#)). More detailed gradual results are available upon request.

<sup>10</sup> The Digital Atlas of Roman and Medieval Civilizations (DARMC) is a website with free GIS maps for the Roman and medieval worlds (see <http://darmc.harvard.edu/icb/icb.do>).

**Table 4**  
The effect of rail stations on subcenter formation, (IV) Probit Marginal effects.

Dependent var.:	Probability of being (part of) a subcenter									
	All suburban municipalities							Without always subcenters		
	Number of lines-stations		Number of stations		Dummy = 1 for municipality with stat			Dummy = 1 for municipality with stat		
Variable type:	68-10		68-10		68-10			75-10		
Period/year:	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
<b>Panel A: OLS results</b>										
Station variable	0.068 <sup>a</sup> (0.007)	-0.000 (0.001)	0.069 <sup>a</sup> (0.007)	-0.000 (0.001)	0.141 <sup>a</sup> (0.013)	0.003 (0.004)	0.004 (0.005)	0.004 (0.004)	0.010 (0.007)	0.010 <sup>c</sup> (0.006)
Pseudo R <sup>2</sup>	0.19	0.88	0.18	0.88	0.17	0.88	0.87	0.81	0.88	0.85
<b>Panel B: TSLS results</b>										
Station variable	0.087 <sup>a</sup> (0.024)	0.027 <sup>b</sup> (0.012)	0.092 <sup>a</sup> (0.025)	0.028 <sup>b</sup> (0.013)	0.142 <sup>a</sup> (0.046)	0.040 <sup>c</sup> (0.022)	0.047 <sup>c</sup> (0.025)	0.053 <sup>b</sup> (0.022)	0.035 <sup>c</sup> (0.021)	0.075 <sup>c</sup> (0.045)
F-S statistic	73.24	26.99	71.68	25.34	127.12	41.11	39.41	38.04	33.81	31.63
Instrument:	Dummy = 1 if municipality is crossed by a 1870 rail									
ln(Densities)	N	Y	N	Y	N	Y	Y	Y	Y	Y
ln(Dist to CBD)	N	Y	N	Y	N	Y	Y	Y	Y	Y
Geography	N	Y	N	Y	N	Y	Y	Y	Y	Y
History	N	Y	N	Y	N	Y	Y	Y	Y	Y
Socioeconomy	N	Y	N	Y	N	Y	Y	Y	Y	Y
Observations:	7680	7680	7680	7680	7680	7680	6400	6115	1223	1223
		(1280 suburban municipalities × 6 census years)					(1280 × 5)		(1223 × 5)	

Notes: Regressions in columns 1 to 8 include year effects. Robust standard errors and are in parentheses (and are clustered by municipality in regressions in columns 1 to 8). <sup>a</sup>, <sup>b</sup>, and <sup>c</sup> indicates significant at 1, 5, and 10 percent level, respectively.

While OLS conditional results in Panel A columns 2 and 4 show non-significant zero marginal effects, TSLS result in Panel B column 4 indicates that an additional station increases the probability that the municipality becomes (part of) a subcenter by 2.8%. This effect is exactly the same as that of having an additional line stopping in the municipality (Panel B column 2). This does not come as a surprise given that most of the suburban municipalities are only crossed by one train line, so that the number of lines-stations is actually very close to the number of stations.

Regarding the existence of a station, we estimate the same non-significant zero OLS effect (Panel A column 6), and a slightly larger TSLS effect, around 4% over the whole period (Panel B column 6). This TSLS effect increases to 4.7% when we focus on the 1975–2010 period (Panel B column 7), suggesting either that the effect is delayed in time, or that the transportation system built after 1975 (mostly the RER) explains a larger part of the overall effect.

In order to dig further into this time variation, we focus on the 1975–2010 period in columns 8 to 10, taking the municipalities systematically identified as subcenters out of the sample. While most of the estimated OLS marginal effects remain non-significant, the TSLS estimated effect over the period jumps to 5.3% (Panel B column 8) confirming the idea of an reinforced effect in the most recent period (the effect goes from 3.5% in 1975 to 7.5% in 2010 (Panel B columns 9 and 10), but the difference is not significant).

We check the robustness of our results in Appendix C. In Table C.1 Panel A we show that estimates are robust to subcenter size and definition: the effect is always between 3.5% and 4.4%, whether we focus on subcenters with more or less than 50,000 jobs (columns 1 and 2), and whether we rely on subcenters identified using the 5% criterion (columns 3 and 4) (instead of 10% in the main results). In Table C.2 Panel A we test the validity of our identification strategy by following Chandra and Thompson (2000) and dropping some observations. Since the 1870 railroad network was probably planned to serve the most important municipalities during the 19th centuries, we first drop municipalities that were important. We do not have population data at the municipality level for these years, as a result we use our historical

dummy variables that signal the most important towns through history. That is, we drop municipalities that were Roman settlements and/or major towns during the 10th and 19th centuries and/or with a monastery built between the 12th and 16th centuries (columns 1 and 2). Alternatively, we also drop observations of municipalities with a rail station built during the 19th century (columns 3 and 4). In both cases, results still show a significant and positive effect of having a rail station on the probability of becoming (part of) a subcenter.

We now refine our results by investigating the train station effect, looking alternatively at two different train types: suburban train vs. RER. The corresponding results are displayed in columns 1 to 4 of Table 5 for the former train type, and in columns 5 to 8 for the latter. Again, OLS and TSLS results are in Panels A and B, respectively. Over the 1968–2010 period, we estimate that the presence of a suburban train station in a municipality (column 2) increases the likelihood that it becomes a subcenter by 0.8% (OLS Panel A) and 3.4% (TSLS Panel B). As before, this effect slightly increases (to 0.9–3.9%) when we focus on the 1975–2010 period (column 3), and to 0.7–4.4% once we exclude municipalities that are always identified as a subcenter (column 4).

On the other hand, OLS and TSLS results for the RER (columns 5 to 8) show significant but opposite effects: while all conditional OLS marginal effects are negative, TSLS marginal effects are positive. In particular, the TSLS results for the RER reveal that this particular type of train has a much stronger impact on subcenter formation than the suburban train. The existence of a RER station is indeed found to increase the probability of becoming (part of) a subcenter by 14% over the 1968–2010 period (Panel B column 6), an effect about four times as large as for suburban trains. Interestingly, looking at the later period (after 1975) does not show a significantly different effect (13.5%, Panel B column 7).

The difference between our TSLS estimates and their OLS counterparts suggests that construction of railroads in Paris is endogenous. Why? It may be due to classical measurement error, but, since similar TSLS-OLS differences are found when using different measures of railroads (number of lines-stations vs. number of stations vs. dummy for municipalities with stations), we rule out this possibility.

**Table 5**  
The effect of train and RER stations on subcenter formation, (IV) Probit Marginal effects.

Dependent var.:	Probability of being (part of) a subcenter							
	Train stations				RER stations			
	All suburban municipalities		Without always sub		All suburban municipalities		Without always sub	
	68-10 [1]	68-10 [2]	75-10 [3]	75-10 [4]	68-10 [5]	68-10 [6]	75-10 [7]	75-10 [8]
<b>Panel A: OLS results</b>								
Station dummy	0.100 <sup>a</sup> (0.014)	0.008 <sup>b</sup> (0.004)	0.009 <sup>b</sup> (0.004)	0.007 <sup>b</sup> (0.003)	0.123 <sup>a</sup> (0.017)	-0.012 <sup>a</sup> (0.003)	-0.012 <sup>a</sup> (0.004)	-0.006 <sup>b</sup> (0.003)
Pseudo R <sup>2</sup>	0.06	0.88	0.87	0.81	0.04	0.88	0.88	0.81
<b>Panel B: TSLS results</b>								
Station dummy	0.175 <sup>a</sup> (0.049)	0.034 <sup>c</sup> (0.020)	0.039 <sup>c</sup> (0.021)	0.044 <sup>a</sup> (0.017)	0.680 <sup>a</sup> (0.191)	0.140 <sup>a</sup> (0.046)	0.135 <sup>a</sup> (0.036)	0.102 <sup>a</sup> (0.034)
F-S statistic	114.12	44.56	43.62	36.83	29.43	10.74	9.91	11.21
Instrument:	Dummy = 1 if municipality is crossed by a 1870 railroad line							
ln(Densities)	N	Y	Y	Y	N	Y	Y	Y
ln(Dist to CBD)	N	Y	Y	Y	N	Y	Y	Y
Geography	N	Y	Y	Y	N	Y	Y	Y
History	N	Y	Y	Y	N	Y	Y	Y
Socioeconomy	N	Y	Y	Y	N	Y	Y	Y
Observations:	7680 (1280 × 6 years)	7680	6400 (1280 × 5)	6115 (1223 × 5)	7680 (1280 × 6 years)	7680	6400 (1280 × 5)	6115 (1223 × 5)

Notes: All regressions include year effects. Robust standard errors are clustered by municipality and are in parentheses. <sup>a</sup>, <sup>b</sup>, and <sup>c</sup> indicates significant at 1, 5, and 10 percent level, respectively.

It may also be due to a negative correlation between the existence of a station and the error term because of missing variables or reverse causation. Despite controlling for geography, population history, socioeconomy and year fixed-effects, the possibility remains that the TSLS-OLS differences could be explained by a missing variable such as the municipal land use regulations, which could be associated with a higher probability of becoming (part of) a subcenter and with not having a railroad station. Alternatively, it may be that conditional on controls, less dense non-central municipalities on average experience positive shocks and receive a station. Although not reported for reasons of space, first-stage results confirm this through a negative estimated marginal effect of -0.011 for the log of job density (suburban train probits) and of -0.025 for the log of population density (RER probits).

4.2. Does proximity to rail stations lead to subcenters?

We now want to examine the effect of the distance to a train station on subcenter formation. This presents a double advantage: it enables us to measure the spatial effect of the presence of a train station, and allows to consider the effect of a train station on municipalities that do not possess any. In other words, we investigate whether the effect of a rail station can go beyond the boundaries of the municipality where the station is located.

To this aim, we now use the distance (in log) of a municipality's centroid to the closest train station as the *train station variable*. Table 6 reports OLS and TSLS results in Panels A and B, respectively. Columns 1 to 4 present results for all railroad stations. In columns 5 to 8, we focus on suburban train stations. Finally, we analyze the effect of proximity to RER stations in columns 9 to 11.

The main conclusion to be drawn from the TSLS results (Panel B) is that train stations have extended spatial effects: being closer to a station increases the probability to be (part of) a subcenter, even for municipalities without any station. As in the previous section, we find a very similar effect of the proximity to any type of train or to suburban train alone. In this case, getting closer to a station by one kilometer increases the probability of becoming a subcenter by 2.4% (columns 2

and 6), or by 2.7% considering the 1975-2010 period (columns 3 and 7). On the other hand, the effect of being one kilometer closer to an RER station is estimated at 4.2% (column 10). We obtain our preferred results when we restrict our sample to the 977 suburban municipalities that do not have any station within their boundaries (columns 4, 8 and 11). This effect, of 3% for suburban trains and 4.9% for RER, confirms the spatial effect of railroad stations.

As previously the larger TSLS estimates (vs. OLS) suggest the endogeneity of railroad construction in Paris and could be associated to a missing variable (local land use regulations) and to reverse causation (positive shocks to less dense non-central municipalities).

Finally, we check that our results are robust to subcenter size (more or less than 50,000 jobs) and definition (subcenters identified using the 5% threshold instead of 10%), and, by following Chandra and Thompson (2000), to a more restrictive sample after dropping observations for historical towns and municipalities crossed by railroads during the 19th century. These tests are reported in Table C.1 Panel B and Table C.2 Panel B in Appendix C.

To summarize the results of this section, two main points can be highlighted. First, railroad stations do play a role in the subcenter formation process, and this effect is spatially lagged: the existence of a train station increases the probability of becoming part of a subcenter by 4 to 5%, and decreases at a rate of about 3% per kilometer. Second, the RER is the type of train having the most important effect, with a direct effect of around 14% for municipalities with a station, and a spatial decay of about 5% per kilometer.

5. Conclusions

In this paper we investigate the effect of railroad construction on the emergence of employment subcenters in metropolitan Paris between 1968 and 2010. Because of the potential endogeneity problem of railroad provision, we rely on IV estimations that use historical instruments built on the 19th century railroad network: a dummy for municipalities crossed by 1870 railroads, and the (log) distance to the nearest 1870 railroad.

**Table 6**  
The effect of rail proximity on subcenter formation, (IV) Probit Marginal effects.

Dependent var.:	Probability of being (part of) a subcenter										
	All suburban stations				Train stations				RER stations		
	All suburban muni		Without muni-stat		All suburban muni		Without muni-stat		All muni	Without muni-stat	
	68-10	68-10	75-10	75-10	68-10	68-10	75-10	75-10	75-10	75-10	75-10
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]
<b>Panel A: OLS results</b>											
ln(Distance)	-0.080 <sup>a</sup>	-0.006 <sup>c</sup>	-0.008 <sup>b</sup>	-0.003	-0.065 <sup>a</sup>	-0.008 <sup>a</sup>	-0.009 <sup>a</sup>	-0.007 <sup>a</sup>	-0.049 <sup>a</sup>	0.011 <sup>a</sup>	0.007 <sup>a</sup>
	(0.007)	(0.003)	(0.003)	(0.003)	(0.007)	(0.002)	(0.003)	(0.001)	(0.005)	(0.002)	(0.001)
Pseudo R <sup>2</sup>	0.26	0.88	0.88	0.94	0.14	0.88	0.88	0.95	0.13	0.89	0.96
<b>Panel B: TSLS results</b>											
ln(Distance)	-0.092 <sup>a</sup>	-0.024 <sup>a</sup>	-0.027 <sup>a</sup>	-0.027 <sup>a</sup>	-0.105 <sup>a</sup>	-0.024 <sup>a</sup>	-0.027 <sup>a</sup>	-0.030 <sup>b</sup>	-0.127 <sup>a</sup>	-0.042 <sup>b</sup>	-0.049 <sup>a</sup>
	(0.014)	(0.007)	(0.007)	(0.007)	(0.016)	(0.009)	(0.009)	(0.012)	(0.017)	(0.019)	(0.019)
F-S statistic	203.20	82.82	82.44	39.92	171.24	43.08	42.40	19.25	87.80	21.28	12.92
Instrument:	ln(Distance to the nearest 1870 railroad line)										
ln(Densities)	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y
ln(Dist to CBD)	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y
Geography	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y
History	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y
Socioeconomy	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y
Observations:	7680	7680	6400	4885	7680	7680	6400	4885	6400	6400	4885
	(1280 × 6)		(1280 × 5)	(977 × 5)		(1280 × 6)	(1280 × 5)	(977 × 5)		(1280 × 5)	(977 × 5)

Notes: All regressions include year effects. Robust standard errors are clustered by municipality and are in parentheses. <sup>a</sup>, <sup>b</sup>, and <sup>c</sup> indicates significant at 1, 5, and 10 percent level, respectively.

After briefly describing Paris’ job location main characteristics and its rail transit, we focus on the suburbs to study the spatial pattern of Paris’ job decentralization. We identify employment subcenters and, despite some municipalities emerge as (part of) subcenters whereas others were dropped, the number of subcenters grew from 21 in 1968 to 35 in 2010. Since employment growth in these subcenters was very intense over the period, we conclude that employment decentralization in Paris is more clustered around subcenters than diffuse, thus reinforcing the polycentric nature of the city.

Finally, we investigate whether railroads cause the emergence of employment subcenters in Paris and our results confirm the causal effect. On average, the probability of a suburban municipality of being (part of a) subcenter increases by 5% if the municipality has a rail station within its boundaries, or by 3% if, although not having a rail station, municipality proximity to a rail station increases by 10%.

Results for the RER confirm its effect is stronger: a 10% and a 5% increase in the probability of becoming (part of a) subcenter if the municipality has a RER station or if municipality proximity to a RER station increases by 10%, respectively.

The contribution of the paper is relevant because, as far as we know, it provides the first empirical evidence on the causal effect of transportation (railroad) on subcenter formation. Furthermore, these new results are useful for urban planners facing and dealing the consequences of urban growth and, in particular, of population suburbanization, employment decentralization and urban sprawl: while these phenomena might reduce city’s agglomeration economies (Glaeser and Kahn, 2004), the emergence of employment subcenters can potentially compensate and even overcome these losses by offering new agglomeration economies and avoiding CBD’s congestion costs (McMillen, 2004).

### Appendix A. Rail transit in Paris: Past and present

One of the main purposes of this paper is to evaluate whether and to what extent transportation has fostered the emergence of employment subcenters in metropolitan Paris. However, first we need to deal with an identification issue because transportation and its improvements are not placed randomly. On the contrary, they are endogenous to employment and/or population location and growth. Planners may for instance decide to improve the connection of deprived areas in order to boost their economic activity or attract population. In order to address this issue, we adopt an instrumental variable approach in which some variables, named instruments, are used as sources of exogenous variation for our transportation endogenous variables.

Recent literature highlights the advantages in terms of exogeneity and relevance of using ‘historical’ and ‘planned’ instruments. For instance, Baum-Snow (2007); Michaels (2008) and Duranton and Matthew (2012) use the 1947 plan of the interstate highway system as an instrument for modern highways in the US, and Duranton and Matthew (2012) additionally rely on the 1898 railroad network. García-López (2012) uses the ancient Roman roads, and the 19th century main road and railroad networks as instruments for highways and railroads in metropolitan Barcelona. Finally, García-López et al. (2015) use the ancient Roman roads and the 1760 Bourbon roads (post routes) to instrument current highways in Spain.

Following the above mentioned literature, we instrument modern railroads in metropolitan Paris with a historical instrument, the 1870 railroad network. The first French railroads were built at the beginning of the 19th century, but slightly later than in the UK due to Napoleon wars: the first line connecting Paris to a city located 18 km away (Saint-Germain) was not opened before 1837. In 1870, the railroad network was based on 698 km of railroad lines. Due to the high levels of centralization in France, it had a star-shaped form centered around Paris (Fig. A.1).

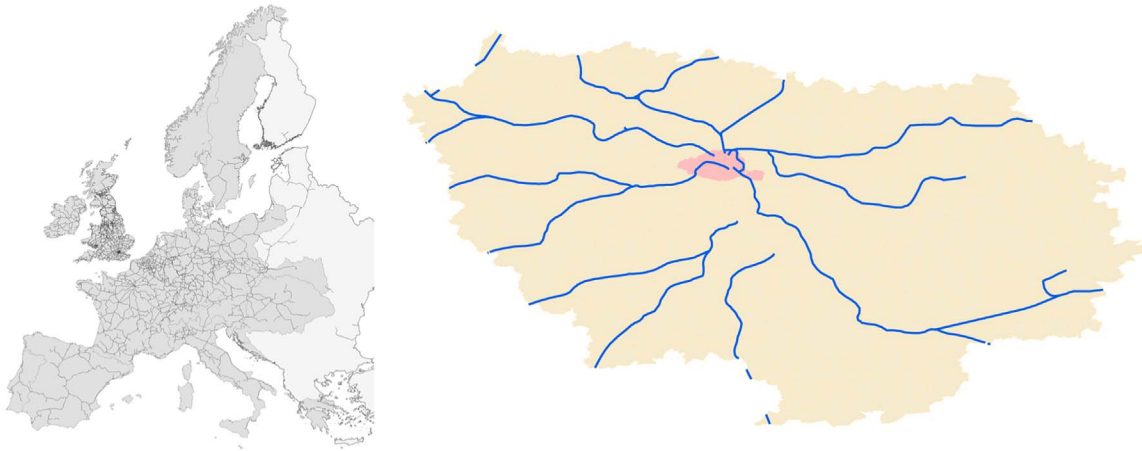


Fig. A.1. The 1870 railroads. Source: Own elaboration based on Marti-Henneberg (2013) maps.

*Is the 1870 rail network a valid instrument?*

As above mentioned, the fact that modern roads and railroads were built following ancient infrastructures has already been argued and used in the literature. Common sense would suggest that in France as well, past infrastructures shape current ones due to practical reasons: it is easier and cheaper to build new transportation infrastructures as an improvement of old ones for instance, or close to them (Duranton and Matthew, 2012). We now test empirically the credibility of this assumption in the context of the metropolitan area of Paris. To do so, we conditionally regress our endogenous rail variables on their historical counterparts and some control variables:

$$2010 \text{ Rail transit variable} = \alpha_0 + \alpha_1 \times 1870 \text{ rail transit variable} + \sum_i (\alpha_{2,i} \times \text{control variables}) \tag{A.1}$$

It is important to point out the importance of the control variables, in particular geography and history. Although ancient transportation infrastructures may be exogenous because of the length of time since they were built, the significant changes undergone by society and economy in the intervening years, and, in particular, because neither of them were built to anticipate employment and population changes in a distant future; it is also true that other factors such as the geography are likely to have influenced the construction and location of both ancient and modern transportation infrastructures for obvious reasons related to the feasibility and convenience of infrastructure building. From this point of view, it is crucial to include geographic characteristics such as altitude, index of terrain ruggedness, and elevation range as controls to comply with the exogeneity condition.

On the other hand, it is equally important to control for the historical context, since it may explain both the presence of former infrastructure and the economic importance of present-days municipalities. In order to fulfill the exclusion restriction, and because there are no historical employment and population data at the municipal level prior to 1962 and 1968, we control for history by including the

**Table A.1**  
Modern rail transit as a function of past rail transit, OLS estimates.

Dependent var.:	Panel A: Rail stations					Dependent var.:	Panel B: Proximity to rail stations		
	Dummy = 1 if muni with station						ln(Dist to nearest station)		
Rail type:		Rail	Train	RER		Rail	Train	RER	
Period/year:	75-10	1975	2010	75-10	75-10	Period:	75-10	75-10	75-10
	[1]	[2]	[3]	[4]	[5]		[6]	[7]	[8]
Dummy = 1 if crossed by 1870 rail	0.179 <sup>a</sup> (0.029)	0.177 <sup>a</sup> (0.030)	0.168 <sup>a</sup> (0.030)	0.173 <sup>a</sup> (0.028)	0.031 <sup>a</sup> (0.009)	ln(Distance to nearest 1870 rail)	0.180 <sup>a</sup> (0.029)	0.121 <sup>a</sup> (0.028)	0.093 <sup>a</sup> (0.026)
Adjusted R <sup>2</sup>	0.37	0.30	0.39	0.24	0.24	Adjusted R <sup>2</sup>	0.54	0.42	0.71
F-S statistic	38.04	33.81	31.63	36.83	11.21	F-S statistic	39.92	19.25	12.92
ln(Densities)	Y	Y	Y	Y	Y	ln(Densities)	Y	Y	Y
ln(Dist to CBD)	Y	Y	Y	Y	Y	ln(Dist to CBD)	Y	Y	Y
Geography	Y	Y	Y	Y	Y	Geography	Y	Y	Y
History	Y	Y	Y	Y	Y	History	Y	Y	Y
Socioeconomy	Y	Y	Y	Y	Y	Socioeconomy	Y	Y	Y
Observations:	6115 (1223 × 5)	1223	1223	6115 (1223 × 5)	6115	Observations:	4885	4885	4885 (977 muni × 5 years)

Notes: Pooled regressions in Columns 1 and 4 to 8 include year effects. Cross section regressions in Columns 2 and 3 include a constant. Columns 1 to 3, Columns 4 and 5, and Columns 6 to 8 show first-stage results for regressions in Table 4 Columns 8 to 10, Table 5 Columns 4 and 8, and Table 6 Columns 4, 8 and 11, respectively. Robust standard errors are clustered by municipality and are in parentheses. <sup>a</sup>, <sup>b</sup>, and <sup>c</sup> indicates significant at 1, 5, and 10 percent level, respectively.

population level in 1962 and dummy variables indicating (1) whether municipalities were Roman settlements, (2) whether they used to be major towns between the 10th and the 15th centuries and (3) between the 16th and the 19th centuries, and (4) whether they had a monastery built between the 12th and 16th centuries. These dummy variables come from the Digital Atlas of Roman and Medieval Civilizations, with the exception of the major cities of the 16th to 19th centuries which are identified in [Bairoch \(1988\)](#). To put it differently, we assume *conditional* exogeneity of the proposed instruments, as suggested by ([Duranton and Matthew, 2012](#)).

Finally, it is important to notice that, to satisfy the exclusion restriction, our endogenous variables and potential instruments are not exactly built in the same way. While the former refer to stations (dummy for municipalities with modern station/s, distance to the nearest modern station) the latter refer to railroad lines (dummy for municipalities crossed by 1870 railroad line/s, distance to the nearest 1870 railroad line). Furthermore, in order to support the validity of our identification strategy we run some robustness checks in [Appendix C Table C.2](#). Following [Chandra and Thompson \(2000\)](#), we exclude historical towns and municipalities with 1870 stations/lines and restrict our attention to municipalities that were ‘accidentally’ crossed by 1870 railroads and did not received 1870 stations.

Regarding the relevance of our potential instruments, [Table A.1](#) shows results for versions of Eq. (A.1) in which we analyze the relationship between modern and past railroads in terms of the presence of stations and proximity to them. In particular, in Panel A, we study whether suburban municipalities crossed by a 1870 rail receive a rail stations. In all cases (pooled vs. cross-section regressions in columns 1 and 2-3, all railroads vs. train and RER regressions in columns 1 and 4-5) we find significant and positive coefficients for the presence of 1870 rails. In Panel B, we estimate the effect of municipality proximity to 1870 rail on the municipality proximity to the nearest modern rail. Conditional on control variables, estimated coefficients for the 1870 distance variable are positive and highly significant. As a whole, results in [Table A.1](#) clearly show that historical rails matter for modern rail construction and location.

**Appendix B. LWR and urban spatial structure in metropolitan Paris, 1968–2010**

**Table B.1**  
Employment spatial structure and LWR: A benchmark to identify subcenters.

	1968	1975	1982	1990	1999	2010
1% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
3% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
5% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
7% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
9% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
10% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
30% LWR Benchmark	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric	Polycentric
50% LWR Benchmark	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric
70% LWR Benchmark	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric
90% LWR Benchmark	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric	Monocentric

**Table B.2**  
Employment spatial structure and LWR: Akaike information criterion.

	1968	1975	1982	1990	1999	2010
1% LWR Akaike inf. crit.	746	767	770	778	774	782
3% LWR Akaike inf. crit.	397	433	446	462	460	479
5% LWR Akaike inf. crit.	346	386	402	423	422	446
7% LWR Akaike inf. crit.	336	377	393	417	417	444
9% LWR Akaike inf. crit.	340	380	396	421	421	450
10% LWR Akaike inf. crit.	345	386	401	426	426	456
30% LWR Akaike inf. crit.	573	598	593	605	590	631
50% LWR Akaike inf. crit.	817	837	820	828	797	843
70% LWR Akaike inf. crit.	1081	1112	1094	1109	1067	1120
90% LWR Akaike inf. crit.	1284	1331	1318	1346	1301	1364

**Appendix C. Does rail transit cause subcenter formation? Robustness checks**

**Table C.1**

The effect of rail on subcenter formation, IV Probit - Marginal effects: Robustness to subcenter size and significance.

Panel A: The effect of rail stations					Panel B: The effect of proximity to rail stations				
Dependent var.:	Probability of being subcenter				Dependent var.:	Probability of being subcenter			
	Subcenter jobs		5% residuals			Subcenter jobs		5% residuals	
	≥50, 000 75-10 [1]	<50, 000 75-10 [2]	All obs 75-10 [3]	Without alw-sub 75-10 [4]		≥50, 000 75-10 [5]	<50, 000 75-10 [6]	All obs 75-10 [7]	Without alw-sub 75-10 [8]
Station dummy	0.036 <sup>c</sup> (0.020)	0.039 <sup>c</sup> (0.023)	0.035 <sup>c</sup> (0.021)	0.044 <sup>b</sup> (0.021)	ln(Distance)	-0.015 <sup>a</sup> (0.005)	-0.017 <sup>a</sup> (0.006)	-0.013 <sup>c</sup> (0.008)	-0.011 <sup>c</sup> (0.006)
F-S statistic	34.20	42.76	38.95	38.29	F-S statistic	81.35	80.63	80.41	84.55
ln(Densities)	Y	Y	Y	Y	ln(Densities)	Y	Y	Y	Y
ln(Dist to CBD)	Y	Y	Y	Y	ln(Dist to CBD)	Y	Y	Y	Y
Geography	Y	Y	Y	Y	Geography	Y	Y	Y	Y
History	Y	Y	Y	Y	History	Y	Y	Y	Y
Socioeconomy	Y	Y	Y	Y	Socioeconomy	Y	Y	Y	Y
Observations:	6214	6117	6400	6195	Observations:	6214	6117	6400	6195
Instrument:	Dummy = 1 if crossed by a 1870 rail				Instrument:	ln(Dist to the nearest 1870 rail)			

Notes: All regressions include year effects. Robust standard errors are clustered by municipality and are in parentheses. <sup>a</sup>, <sup>b</sup>, and <sup>c</sup> indicates significant at 1, 5, and 10 percent level, respectively.

**Table C.2**

The effect of rail on subcenter formation, IV Probit - Marginal effects: Robustness to identification strategy.

Panel A: The effect of rail stations					Panel B: The effect of proximity to rail stations				
Dependent var.:	Probability of being subcenter				Dependent var.:	Probability of being subcenter			
	No historic towns		No 1870 stations			No historic towns		No 1870 lines	
	All obs 75-10 [1]	Without alw-sub 75-10 [2]	All obs 75-10 [3]	Without alw-sub 75-10 [4]		Without mun-stat 75-10 [5]	Without alw-sub 75-10 [6]	Without mun-stat 75-10 [7]	Without alw-sub 75-10 [8]
Station dummy	0.067 <sup>a</sup> (0.023)	0.046 <sup>b</sup> (0.022)	0.049 <sup>b</sup> (0.023)	0.037 <sup>c</sup> (0.021)	ln(Distance)	-0.023 <sup>a</sup> (0.006)	-0.010 <sup>c</sup> (0.005)	-0.025 <sup>a</sup> (0.004)	-0.020 <sup>a</sup> (0.004)
F-S statistic	34.55	35.13	8.53	10.64	F-S statistic	52.85	50.69	95.90	90.45
ln(Densities)	Y	Y	Y	Y	ln(Densities)	Y	Y	Y	Y
ln(Dist to CBD)	Y	Y	Y	Y	ln(Dist to CBD)	Y	Y	Y	Y
Geography	Y	Y	Y	Y	Geography	Y	Y	Y	Y
Lagged pop	Y	Y	Y	Y	Lagged pop	Y	Y	Y	Y
Socioeconomy	Y	Y	Y	Y	Socioeconomy	Y	Y	Y	Y
Observations:	6185	5955	5385	5245	Observations:	4780	4675	4110	3945
Instrument:	Dummy = 1 if crossed by a 1870 rail				Instrument:	ln(Dist to the nearest 1870 rail)			

Notes: All regressions include year effects. Robust standard errors are clustered by municipality and are in parentheses. <sup>a</sup>, <sup>b</sup>, and <sup>c</sup> indicates significant at 1, 5, and 10 percent level, respectively.

**References**

Anas, Alex, Kim, Ikki, 1996. General equilibrium models of polycentric urban land use with endogenous congestion and job agglomeration. *J. Urban Econ.* 40 (2), 232–256.

Arribas-Bel, Daniel, Sanz-Gracia, Fernando, 2014. The validity of the monocentric city model in a polycentric age: us metropolitan areas in 1990, 2000 and 2010. *Urban Geogr.* 35 (7), 980–997.

Bairoch, Paul, 1988. *Cities and Economic Development: from the Dawn Of History to the Present.* University of Chicago Press, Chicago.

Baum-Snow, Nathaniel, Kahn, Matthew E., 2000. The effects of new public projects to expand urban rail transit. *J. Public Econ.* 77 (2), 241–263.

Baum-Snow, Nathaniel and Matthew E. Kahn. 2005. Effects of urban rail transit expansions: Evidence from sixteen cities, 1970–2000. *Brookings-Wharton Papers on Urban Affairs.* 147–206.

Baum-Snow, Nathaniel, Brandt, Loren, Vernon Henderson, J., Turner, Matthew A., Zhang, Qinghua, 2015. Roads, railroads and decentralization of chinese cities. Mimeo.

Baum-Snow, Nathaniel, 2007. Did highways cause suburbanization? *Q. J. Econ.* 122 (2), 775–805.

Baum-Snow, Nathaniel, 2010. Changes in transportation infrastructure and commuting patterns in us metropolitan areas 1960-2000. *Am. Econ. Rev. Pap. Proc.* 100, 378–382.

Baum-Snow, Nathaniel, 2014. Urban transport expansions, employment decentralization, and the spatial scope of agglomeration economies. Mimeo.

Berliant, Marcus, Wang, Ping, 2008. Urban growth and subcenter formation: a trolley ride from the Staples Center to Disneyland and the Rose Bowl. *J. Urban Econ.* 63 (2), 679–693.

Berliant, Marcus, Peng, Shin-Kun, Wang, Ping, 2002. Production externalities and urban configuration. *J. Econ. Theory* 104 (2), 275–303.

Bollinger, Christopher, Ihlanfeldt, Keith, 1997. The impact of rapid rail transit on economic development: the case of atlanta's marta. *J. Urban Econ.* 42 (2), 179–204.

- Bollinger, Christopher, Ihlanfeldt, Keith, 2003. The intraurban spatial distribution of employment: which government interventions make a difference? *J. Urban Econ.* 53 (3), 396–412.
- Bowes, David R., Keith, R.Ihlanfeldt, 2001. Identifying the impacts of rail transit stations on residential property values. *J. Urban Econ.* 50 (1), 1–25.
- Brooks, Leah, Lutz, Byron, 2016. Vestiges of transit: Urban persistence at a micro scale. Working paper.
- Chandra, Amitabh, Thompson, Eric, 2000. Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Reg. Sci. Urban Econ.* 30 (4), 457–490.
- Chen, Yihsu, Whalley, Alexander, 2012. Green infrastructure: the effects of urban rail transit on air quality. *Am. Econ. J.: Econ. Policy* 4 (1), 58–97.
- Craig, Steven, Pin, T.Ng, 2001. Using quantile smoothing splines to identify employment subcenters in a multicentric urban area. *J. Urban Econ.* 49 (1), 100–120.
- Craig, Steven G., Janet, E.Kohlhase, Adam, W.Perdue, 2016. Empirical polycentricity: the complex relationship between employment centers. *J. Reg. Sci.* 56 (1), 25–52.
- Diao, Mi, Zhu, Yi, Jiren, Zhu, 2017. Intra-city access to inter-city transport nodes: The implications of high-speed-rail station locations for the urban development of Chinese cities. *Urban Stud.* 54 (10), 2249–2267.
- Duranton, Gilles, Matthew, A.Turner, 2012. Urban growth and transportation. *Rev. Econ. Stud.* 79 (4), 1407–1440.
- Duranton, Gilles, Diego, Puga, 2015. Urban land use. In: Duranton, Gilles, Vernon Henderson, J., Strange, William C. (Eds.), *Handbook of Regional and Urban Economics Volume 5*. North-Holland, Amsterdam, 467–560.
- Fujita, Masahisa, Ogawa, Hideaki, 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. *Reg. Sci. Urban Econ.* 12 (2), 161–196.
- García-López, Miquel-Àngel, Muñoz, Ivan, 2013. Urban spatial structure, agglomeration economies, and economic growth in barcelona: an intra-metropolitan perspective. *Pap. Reg. Sci.* 92 (3), 515–534.
- García-López, Miquel-Àngel, Solé-Ollé, Albert, Viladecans-Marsal, Elisabet, 2015. Does zoning follow highways? *Reg. Sci. Urban Econ.* 53(C), 148–155.
- García-López, Miquel-Àngel, Holl, Adelheid, Viladecans-Marsal, Elisabet, 2015. Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities. *J. Urban Econ.* 85 (0), 52–67.
- García-López, Miquel-Àngel, Hémet, Camille, Viladecans-Marsal, Elisabet, 2017. How does transportation shape intrametropolitan growth? An answer from the Regional Express Rail. *Journal of Regional Science*, Forthcoming.
- García-López, Miquel-Àngel, 2010. Population suburbanization in Barcelona, 1991–2005: Is its spatial structure changing? *J. Hous. Econ.* 19 (2), 119–132.
- García-López, Miquel-Àngel, 2012. Urban spatial structure, suburbanization and transportation in Barcelona. *J. Urban Econ.* 72 (2–3), 176–190.
- Gibbons, Stephen, Machin, Stephen, 2005. Valuing rail access using transport innovations. *J. Urban Econ.* 57 (1), 148–169.
- Giuliano, Genevieve, Small, Kenneth A., 1991. Subcenters in the Los Angeles region. *Reg. Sci. Urban Econ.* 21 (2), 163–182.
- Giuliano, Genevieve, Small, Kenneth A., 1999. The determinants of growth of subcenters. *J. Transp. Geogr.* 7, 189–201.
- Giuliano, Genevieve, Redfeare, Christian L., Agarwal, Ajay, Li, Chen, Zhuang, Duan, 2007. Employment concentrations in Los Angeles, 1980–2000. *Environ. Plann. A* 39 (12), 2935–2957.
- Giuliano, Genevieve, Redfeare, Christian L., Agarwal, Ajay, He, Sylvia, 2012. Network accessibility and employment centers. *Urban Stud.* 49 (1), 77–95.
- Glaeser, Edward L., Matthew E., Kahn, 2001. Decentralized employment and the transformation of the American city. In: Gale, W.G., Pack, J.R. (Eds.), *Brookings-Wharton Papers on Urban Affairs: 2001*. Brookings Institution Press, Washington, 1–47.
- Glaeser, Edward L., Kahn, Matthew E., 2004. Sprawl and urban growth. In: Henderson, J. Vernon, Thisse, Jacques-François (Eds.), *Handbook of Regional and Urban Economics Volume 4*. North-Holland, Amsterdam, 2481–2527.
- Grube-Cavers, Annelise, Patterson, Zachary, 2015. Urban rapid rail transit and gentrification in canadian urban centres: A survival analysis approach. *Urban Stud.* 52 (1), 178–194.
- Helsley, Robert W., Arthur, M. Sullivan, 1991. Urban subcenter formation. *Reg. Sci. Urban Econ.* 21 (2), 255–275.
- Henderson, Vernon, Mitra, Arindam, 1996. The new urban landscape: developers and edge cities. *Reg. Sci. Urban Econ.* 26 (6), 613–643.
- Henderson, Vernon, Slade, Eric, 1996. Development games in non-monocentric cities. *J. Urban Econ.* 34 (2), 207–229.
- Hurst, Needham B., West, Sarah E., 2014. Public transit and urban redevelopment: the effect of light rail transit on land use in Minneapolis, Minnesota. *Reg. Sci. Urban Econ.* 46(C), 57–72.
- Li, Shanjun, Yang, Jun, Qin, Ping, Chonabayashi, Shun, 2017. Wheels of fortune: Subway expansion and property values in Beijing. *J. Reg. Sci.* 56, 792–813.
- Lucas, Robert E., Rossi-Hansberg, Esteban, 2002. On the internal structure of cities. *Econometrica* 70 (4), 1445–1476.
- Marti-Henneberg, Jordi, 2013. European integration and national models for railway networks (1840–2010). *J. Transp. Geogr.* 26 (0), 126–138.
- Mayer, Thierry, Trévién, Corentin, 2015. The impacts of urban public transportation. Evidence from the Paris metropolitan area. Working Paper g2015/03, INSEE.
- McDonald, John F., Paul, J.Prather, 1994. Suburban employment centers: the case of Chicago. *Urban Stud.* 31 (2), 201–218.
- McDonald, John F., 1987. The identification of urban employment subcenters. *J. Urban Econ.* 21 (2), 242–258.
- McMillen, Daniel P., McDonald, John F., 1998. Suburban subcenters and employment density in metropolitan Chicago. *J. Urban Econ.* 43 (2), 157–180.
- McMillen, Daniel P., McDonald, John, 2004. Reaction of house prices to a new rapid transit line: Chicago's Midway Line, 1983–1999. *Real Estate Econ.* 32 (3), 463–486.
- McMillen, Daniel P., Redfeare, Christian L., 2010. Estimation and hypothesis testing for nonparametric hedonic house price functions. *J. Reg. Sci.* 50 (3), 712–733.
- McMillen, Daniel P., Stefani, C.Smith, 2003. The number of subcenters in large urban areas. *J. Urban Econ.* 53 (3), 321–338.
- McMillen, Daniel P., 2001. Nonparametric employment subcenter identification. *J. Urban Econ.* 50 (3), 448–473.
- McMillen, Daniel P., 2003. Identifying sub-centres using contiguity matrices. *Urban Stud.* 40 (1), 57–69.
- McMillen, Daniel P., 2004. Employment densities, spatial autocorrelation, and subcenters in large metropolitan areas. *J. Reg. Sci.* 44 (2), 225–244.
- Michaels, Guy, 2008. The effect of trade on the demand for skill: evidence from the interstate highway system. *Rev. Econ. Stat.* 90 (4), 683–701.
- Muñiz, Ivan, Galindo, Anna, García-López, Miquel-Àngel, 2003. Cubic spline population density functions and satellite city delimitation: the case of Barcelona. *Urban Stud.* 40 (7), 1303–1321.
- Muñiz, Ivan, García-López, Miquel-Àngel, Galindo, Anna, 2008. The effect of employment sub-centres on population density in Barcelona. *Urban Stud.* 45 (3), 627–649.
- OECD, 2017. The Governance of Land use in France: Case studies of Clermont-Ferrand and Nantes Saint-Nazaire. OECD Publishing, Paris.
- Redfeare, Christian L., 2007. The topography of metropolitan employment: identifying centers of employment in a polycentric urban area. *J. Urban Econ.* 61 (3), 519–541.
- Redfeare, Christian L., 2009a. Persistence in urban form: the long-run durability of employment centers in metropolitan areas. *Reg. Sci. Urban Econ.* 39 (2), 224–232.
- Redfeare, Christian L., 2009b. How informative are average effects? Hedonic regression and amenity capitalization in complex urban housing markets. *Reg. Sci. Urban Econ.* 39 (3), 297–306.
- Ryan, Sherry, 2005. The value of access to highways and light rail transit: evidence for industrial and office firms. *Urban Stud.* 42 (4), 751–764.
- Schuetz, Jenny, 2015. Do rail transit stations encourage neighbourhood retail activity? *Urban Stud.* 52 (14), 2699–2723.

