HAL
open science

Do agglomeration economies are lower for polluting sectors?

Emmanuelle Leturque, Mathieu Sanch-Maritan

To cite this version:

Emmanuelle Leturque, Mathieu Sanch-Maritan. Do agglomeration economies are lower for polluting sectors?. 2019. hal-02347595

HAL Id: hal-02347595

https://hal.science/hal-02347595v1

Preprint submitted on 5 Nov 2019

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.

Do agglomeration economies are lower for polluting sectors?

Emmanuelle Leturque[1] and Mathieu Sanch-Maritan [2,3]

[1]EA LEDi , Université de Bourgogne Franche-Comté

[2]EA CREAM, Université de Rouen-Normandie

[3]UMR CESAER, AgroSup, INRA, Université de Bourgogne Franche-Comté

Abstract

This article explore how the relation between productivity and local city-size can be mitigated by pollution. More specifically, we estimate agglomeration economies considering a new source of heterogeneity among industries: the degree of pollution. Due to pollution perception acting as a dispersion force, we expect *net* agglomeration economies to be lower for polluting firms. In fact, polluting firms may anticipate that households and other firms are reluctant to locate near sources of pollution. In this paper, we exploit spatial data on sectoral emissions for a large number of air pollutants. We define a continuous variable of pollution that varies across sectors and employment zones. Our finding are twofold. First we find that agglomeration economies are lower for polluting sectors. Second we find that negative agglomeration are observed for some key pollutant such as carbon dioxide, nitrogen dioxide, lead or sulfur dioxide.

Keywords: Agglomeration economies, Polluting sectors, Negative externalities

Introduction

Quantifying the benefits and spillovers of agglomeration economies has been a major interest in the urban literature over past decades. Rosenthal and Strange (**?**) find that agglomeration economies elasticity for productivity ranges between 0.03 and 0.08: 1% increase of agglomeration implies an increase of 0.03% to 0.08% in labor productivity. The vast majority of the empirical literature on agglomeration economies studies the net impact of density on labor productivity[1]. These studies measure the overall impact of agglomeration on productivity taking into account both positive and negative effects. Positive agglomeration effects can also turn negative above some city size threshold, or induce some accompanying effects, and one cannot say to which extent positive effects offset the negative, as only the total net impact is evaluated (see for example, Combes and Gobillon, **?**).

The aim of this paper is to estimate agglomeration economies considering a new source of heterogeneity among industries: the degree of pollution. In New Economic Geography (NEG) models, spatial equilibrium is more dispersed when the model incorporates some degree of pollution heterogeneity. For instance, Lange and Quaas (**?**) show that interior equilibria (where not all activities are agglomerated in the Core - namely CP equilibrium) are more frequent once the model takes into account the negative effect of industrial pollution on households.

Due to pollution perception acting as a dispersion force, we expect *net* agglomeration economies to be lower for polluting firms. In fact, polluting firms may anticipate that households and other firms are reluctant to locate near sources of pollution. Households are reluctant to locate near polluted areas as it implies a loss of well-being. Several channels can operate among others: bad reputation links with anticipation of environmental risks, noise, bad smell, unpleasant views or health issues. Moreover, since pollution can directly reduce workers' productivity (Graff Zivin and Neidell, **?** ; Hanna

[1]Drut and Mahieux (**?**) are an exception considering they aim to disentangle the real impact of agglomeration on productivity from pollution effect.

and Oliva, **?**), including their own employees' productivity, degree of pollution may play a negative role in firms' location choices. Pollution can also reduce productivity through a negative impact on other factors. For instance, soil quality is detrimental for agricultural productivity and we may expect that farmers are reluctant to locate nearby polluting industrial firms. More generally, all sectors in which firms use natural resources might be affected by pollution.

In this paper, on *air* pollution stemming from *industrial* activities. Air pollution damages household's *health*, with for instance some clear results in Chay and Greenstone (**?**) considering *Total Suspension Particulate* on asthma and other respiratory diseases. Moreover, this kind of pollution is rather well-known by households as a result of forecasting of daily pollution levels and the warnings seen in various media.

One could think that industrial pollution becomes a side issue compared to agriculture and transport pollution. Although industrial emissions have decreased sharply since 1980, industry remains the main emitter for some key types of pollution. For instance, emissions of SO_2 (sulfur dioxide) from the sectors of energy production and manufacturing industry represent 85% of total emissions in France in 2011 (CITEP, **?**). Ozone concentration results from interaction between NO_x (nitrogen oxide) and VOC (volatile organic chemical). For these two types of emissions, manufacturing industry and energy production remain important emitters with 41% of total VOC emissions and 20% of NO_x emissions in France in 2011.

Our study introduces a new measure of heterogeneity affecting magnitude of agglomeration economies across sectors. As we have information on the level of emissions for all pollutants, we define a continuous variable of pollution that varies across sectors and employment zones. We construct a strategy with an interaction term between density and pollution variable. We estimate the impact of pollution on the amplitude of agglomeration economies thanks to such interaction term.

We have the following result: for sectors emitting some key pollutants as carbon dioxide, nitrogen dioxide, lead or sulfur dioxide, agglomeration benefits will be lower as

pollutant emissions increase. Two source of endogeneity have been taken into account in our work. First, we exploit historical and geological instruments to tackle endogeneity issue on density variables. Moreover, sectoral productivity and pollutant emissions might be both linked with omitted variables such as local negative economic shocks. Therefore, we treat endogeneity issues affecting the interaction term coefficient thanks to a shift-share strategy (Bartik, **?**). We construct a predicted local level of emissions exploiting national rates of emission growth uncorrelated with specific local sectoral shocks.

To do so, we exploit a unique IREP ("Registre des émissions polluantes") database rarely used in economic studies (an exception, Caudeville and Rican,**?**). The IREP dataset records polluting plants constrained to declare levels of pollutant emissions, wastes and dangerous inputs. We match polluting emissions and wages from DADS ("Données Annuelles de Déclarations Sociales") database at the sectoral and employment zone level. Hence, we are able to identify wages in polluting sectors at a fine spatial scale over the time period 2003-2013.

The paper is built in the following way: we first present the literature on agglomeration economies and wonder how the effect of industrial pollution is assessed so far. We then detail our identification strategy: we implement a pooled estimation with an interaction term between agglomeration and pollution emission. We test if the impact of industrial atmospheric pollution affects agglomeration economies both for polluting firms. Lastly, we present our results and some guidelines for future research.

1 Literature review

Through the review of literature, we aim to enlighten links between NEG with pollution literature and agglomeration economies evaluation literature. To do so, we first summarize the main results of theoretical contributions on the impact of industrial pollution in a economic geography model. Then, we present methodology and results

of empirical works agglomeration economies. We focus on the best way to introduce polluting emissions in such work based on previous works.

1.1 NEG and industrial polluting emissions

Building on the sample of NEG models explained in paper 2, we start by summarizing the main predictions of the literature regarding our empirical question. Most of the papers cited below consider only two sectors: traditional sector and industrial sector. Industrial production leads to polluting emissions. This pollution may directly impact household's utility and agricultural sector productivity. Several channels operate in this situation. First, when qualified workers (or entrepreneurs) are mobile, they take into account the quantity of polluting emissions by region in their location choice (For instance: Elbers and Withagen, **?**; van Marrewijk, **?**; Lange and Quaas, **?**; Rauscher, **?**). Their behavior depends on the characteristics of pollution (transboundary or not), the source of pollution (e.g. industrial or residential sectors), and the individual parameters (sensitivity toward pollution). When industrial damages are integrated, local pollution is simply measured as the number of polluting firms in the region. Because workers are less disposed to locate in the region with the largest number of industrial firms, market in this region becomes thinner. Yet, market size is one of the main agglomeration forces. If this force decreases while the dispersion forces are still active, firm are less willing to agglomerate in the *Core* region. Similar reasoning can be applied with inter-sectoral impact. Industrial pollution alters traditional sector productivity. In the end, labor productivity impact local wages, so a decrease in wage represents a decrease of purchasing power that deteriorates home market effect (Zeng and Zhao, **?**). Allowing firms to emit pollution can change spatial equilibrium outcome, and then spatial equilibrium becomes more dispersed.

While in previous research cited above, all industrial firms emit pollution, this paper introduces heterogeneity among firms regarding pollution emissions. This strategy

allows the analyze of pollution in a different context than congestion[2], to which it has been previously compared. Wu and Reimer (**?**) suggest a few insights to differentiate congestion from pollution phenomena. Pollution variability across firms or sectors seems much higher than variability in congestion contribution. Moreover, pollution emissions depend on the level of production, which is not evident in the case of congestion. Lastly, pollution may be transboundary while congestion cannot really "transport" itself from a region to a different one.

Wu and Reimer (**?**) build a model with two types of mobile firms: polluting and non-polluting firms. In their case, spatial equilibrium must be computed for both types of firms between two symmetric regions: West and East. If agglomeration occurs, they assume that it is in the East. The model considers the service firms as the non-polluting sector. The following is a summary of their polluting cases with equilibrium outcome:

1. *Clean and Dirty firms have the same levels of productivity.* In this situation, if a dirty firm choose to move to the East, its agglomeration benefits are moderated by the associated pollution brought to the East. That's also why the more pollution is transboundary, the less pollution matters. Agglomeration benefits become higher and this effect dominates. If the polluting sector has the "benchmark" level of productivity, firms relatively less polluting -among the polluting ones- are more concentrated in space.

2. *Clean firms are more productive.* Clean firms will agglomerate, while distribution of the dirty ones depends on the level and mobility of pollution. Indeed, if dirty firms produce relatively low levels of pollution, those firms will also agglomerate. On the contrary, if they produce a high level of pollution, they will better disperse.

3. *Dirty firms are more productive.* Spatial equilibrium will again depend on pollution differential between clean and dirty sectors. When the differential is relatively

[2]Works in economic geography include congestion: **?**, **?**, Fujita and Thisse (**?**).

small, agglomeration effect is stringer, and the equilibrium is an agglomerated one. If the difference is larger (but not too large), dirty firms still agglomerate when clean ones prefer to disperse. Finally, when pollution differential is large enough, the equilibrium of this economy is a dispersed one.

Wu and Reimer (**?**) give intuition about the mechanisms behind spatial equilibrium when polluting firms are distinguished from the clean ones.

Heterogeneity in NEG: can pollution intensity be considered as a source of heterogeneity?

Heterogeneity among locations has been one key issue in economic geography. We displays usual sources of heterogeneity, among which pollutant emissions could figure in. We do so because several forms of heterogeneity in terms of spatial organization have been empirically tested. First, "Macro heterogeneity" among regions is explained by NEG models where mechanisms of agglomeration drive to asymmetric allocations, as CP equilibrium. In the same time, **?** already uses two kinds of workers-qualified and less qualified- to describe agglomeration phenomenon based on qualified labor migration. Then, few papers tackled the question of heterogeneity among agents in theoretical economic geography models. Originally introduced by **?** in a trade framework, firm heterogeneity enlarges deeply the picture of firm interactions in a global world. It provides an explanation of current intra-industry reallocation into international or domestic markets and helps to identify the factors that drive firms to actually enter international markets.

Empirical studies target to determine the size and direction of the effect agglomeration has on local productivity. Firm data availability allows researchers to explain firm productivity and the role of local context. In this sense, empirical findings bring theory to better capture on spatial sorting and selection phenomenon.

Theoretical models take into account diversity of productivity levels among firms.

While identical firm assumption is very convenient in a NEG model, it is also an unrealistic assumption. Firms can vary according to their level of productivity, size, and behavior toward trade. Baldwin and Okubo (**?**) use firm heterogeneity in agglomeration economics, while Behrens and Robert-Nicoud (**?**) make an overview of contributions about heterogeneous agents in an urban setting. Magnitude of agglomeration economies will vary among firms with different productivity levels. The number of competitive firms per industry matters. The following options have been introduced to overcome the limits induced by typical NEG model. First, Dixit-Stiglitz monopolistic competition can be replaced by an oligopolistic framework to introduce firm interactions with quadratic linear utility function (**?**). Second, following **?**, a monopolistic framework with heterogeneity (Baldwin and Okubo, **?**) permits to identify empirical findings. While **?** or Melitz and Ottaviano (**?**) explain that selection effect drives least efficient firms out of international markets or exit the market within their country of origin, Baldwin and Okubo (**?**) focus on spatial selection in some locations. They use firm productivity heterogeneity where the distribution of firm productivity is given for each region. They show that firms from small regions moving to a larger market are the most efficient ones. The firms with largest sales in terms of volume benefit from smaller trade costs when locating in larger markets.

Larger or denser cities (employment areas in our case) experience higher productivity because of local advantages and thanks to agglomeration economies. More productive firms sort in the areas that represent larger markets. Those areas select most efficient firms and the same reasoning can be conducted for workers.

Finally, another source of heterogeneity could be introduced such as pollution intensity, or, more generally, the environmental effort made by each firm. The latter aspect remains hard to evaluate and may need qualitative information on the behavior of entrepreneurs. In an economic geography model, Wu and Reimer (**?**) are the first to introduce two sources of heterogeneity: pollution and productivity. Willingness

to comply with environmental regulation might be correlated with the agglomeration benefits that firms expect.

1.2 Agglomeration economies evaluation with polluting emissions

We want to stress how industrial levels of polluting emission can be introduced in empirical evaluations of agglomeration economies. We first establish the existent links between NEG and his empirical applications. Then, the methodology in agglomeration economies evaluation is displayed and gives grounds for our empirical choices. Finally, we focus on a few relevant empirical studies considering air pollution emissions.

1.2.1 NEG and empirics

Empirical applications of NEG models have been rather sparse (**?**). Empirical studies on NEG models analyze market access measure and plant location choice. However, to go beyond descriptive analysis as concentration indexes or spatial autocorrelation studies with plant location empirics (Redding, **?**; Combes and Gobillon, **?**) are problematic. Direct evaluation of NEG model would require to estimate local determinants of profit, as firms choose their location according to expected local profit. Structural econometric models are able to overcome limits that have been previously raised ; for instance, Head and Mayer (**?**) who study location choice of FDI from Japanese firms in European countries.

If current industrial spatial organization reflects maximization profit behavior, it is interesting to evaluate the importance of agglomeration economies across sectors. Most of the time, papers estimating agglomeration economies that build a theoretical part, construct a model deriving from profit maximization an equation of productivity, depending on local determinants. However, they are quite different from NEG models which relate predominantly to agglomeration determinants rather than to productivity

determinants.

The most direct way to test a theoretical NEG model with pollution would be to identify agglomeration mechanisms taking into account heterogeneous firms according to their amount of polluting emissions. One could, secondly, consider the result, i.e. the local determinants of firm performance with agglomeration as a variable of interest. Our study belongs to the second category. As we remarked it before, the first one remains mostly descriptive[3], while the second has been investigated by Ciccone and Hall (**?**), **?** or **?** for instance, and permits to have guidelines. This approach still gives some intuitions about forces that drive firms to agglomerate and if those who are polluting might benefit in a different way from agglomeration. We test a hypothesis used in theoretical models and an implied aspect contained in theoretical work (for instance, Elbers and Withagen, **?**; Lange and Quaas, **?**; Wu and Reimer, **?**): polluting firms choose less frequently agglomerated areas because they benefit less from agglomeration economies.

1.2.2 Evaluation of agglomeration economies

One of the long-term objectives of quantitative evaluation of static agglomeration economies is to figure out if current city size or activities distribution across space is optimal. To do so, one needs to estimate whether firms and workers have an interest to locate in dense areas through obtaining wages or profit premiums.

The usual procedure is to estimate local productivity toward local determinants including a variable of "agglomeration". More precisely, there are two main possibilities to measure productivity: total factor productivity (*TFP*) or *wage*. Whereas *wage* is an approximation of labor productivity -under perfect competition, *TFP* constitutes a measure of productivity at the firm level. In theory, wage is determined by labor productivity. The multiplicative factor depends on the type of competition setting in

[3]Rosenthal and Strange, (**?**) study the agglomeration determinants.

which the firm belongs to: the higher the monopoly power, the further wage is from labor productivity[4]. The set of covariates contains local determinants such as market access, local industrial composition and, in particular, a measure of "agglomeration", which is most often local employment density Rosenthal and Strange (**?**), computed as the ratio between the number of employees working in the zone and the zone surface.

In most studies following Ciccone and Hall (**?**), the empirical specification uses the logarithm of nominal wage[5] as the outcome and the logarithm of local density as the key explanatory variable. As usually, the log-log form reduces extreme values weight and the density coefficients can be interpreted as an elasticity. The basic estimation equation of agglomeration effects in a reduced form can be summarized in the following way:

$$y_{i,t} = Z_{i,t}\gamma + \eta_{i,t} \ , \tag{1}$$

with $y_{i,t}$ being the logarithm of wage, $Z_{i,t}$ the set of covariates (local productivity determinants including the logarithm of density and labor skills), γ vector of estimated coefficients, $\eta_{i,t}$ random residual.

Depending on the study and the empirical specification, 1% increase of local density implies an increase of productivity from 0.03 to 0.08%. In other words, if the local density is multiplied by 2, it implies an increase of local productivity by $2^{0.08} - 1 \approx 5.7$ % [6] for a coefficient of density equals to 0.08 (upper bond).

Most of empirical studies cited above identify the *causal effect* of local determinants on local productivity. If unobserved variables in the residual are correlated with density, OLS estimator of density will be biased. Since Ciccone and Hall (**?**), researchers

[4]This factor might be correlated to local determinants, using *TFP* permits to bypass this issue. However, using *TFP* requires larger amount of firm data and an estimation of production function for which empirical strategy is still discussed in the literature (Combes and Gobillon **?**)

[5]The choice of nominal wage has been discussed by **?**, controlling for land price does not seem necessary.

[6]As usual, if local density $dens_1$ is multiplied by 2 to achieve $dens_2$, the increase of productivity from w_1 to w_2 is $(2^\beta - 1) \times 100\%$. Indeed, $\ln(\frac{w_2}{w_1}) = \beta \ln(\frac{dens_2}{dens_1}) = \beta \ln(2) = \ln 2^\beta$, we apply exponential function to get: $\frac{w_2}{w_1} = 2^\beta$. It implies that $\frac{w_2 - w_1}{w_1} = 2^\beta - 1$.

have focused on the endogeneity issue. We present our strategy in the next section. We can identify endogeneity issue with the regression equation (1). A positive bias in the estimation of density effect can arise from reverse causality: higher wages -$y_{i,t}$- may attract more workers and consequently increase employment density- $Z_{i,t}$. Then, missing variables may influence both local productivity and explanatory variables including employment density. For instance, infrastructure amenities may create a positive bias if they are not included in the regression because they may simultaneously increase productivity and attract workers. Omitted variables can also generate negative bias: some amenities as cultural ones may not directly influence firm productivity. Because they attract people, land price increases and pushes firm to use less land with respect to labor. This mechanism affects labor productivity negatively. Moreover, endogeneity concerns may arise from individual level. Workers can self-select in some locations according to unobserved individual variables or choose their location according to expected wage they will receive.

There exists a branch of the literature very critical of the way agglomeration economies is evaluated *à la* Ciccone and Hall (**?**). For example, using wage as employment area's productivity has some evident drawbacks: underlying assumption is that through efficient labor market, wage is a good estimation of individual productivity. However, a very few macro territorial elements take into account in such evaluations, and "smaller cities may utilize other attributes and context advantages to boost their locational advantage" (**?**).

1.2.3 Introducing heterogeneity in the specification

The first heterogeneity dimension is the fact that agglomeration economies might not be positive for all cities. Agglomeration phenomenon exists up to a certain level after which dispersion forces become stronger. However, most of empirical studies find a positive impact of density on productivity. Agglomeration economies intensity depends

on the level of agglomeration.

A growing amount of papers tends to introduce heterogeneous effects according to sector characteristics (Brulhart and Mathys, **?**), to firms (**??**), to types of workers (Galeser and Maré,**?**) or to agglomeration size (**?**). Sector[7] heterogeneity has not been so often analyzed in the literature. However, a lot of empirical studies focus on few sectors where agglomeration effects are supposed to be stronger, as manufacturing or R&D for instance. Indeed there exist a large literature on the specificity of R&D sector and the advantages that R&D firms get by locating close to each others. Benefits from innovative activities concentration in R&D has been widely studied (for a literature review, see Carlino and Kerr (**?**)). One of the solutions in the particular case of innovation production is to estimate a production function.

However, building a production function for each sector requires a lot of information and might not be relevant because of large diversity of products. **?** suggest in this case two main possibilities to introduce industry approach in econometric study of agglomeration. First, if agglomeration economies are not equally distributed across sectors, we should find that industries are not spatially concentrated in the same way[8]. Second, agglomeration economies are capitalized and can be observed through labor marginal productivity. We choose to follow this approach and to estimate wages according to local determinants.

Brulhart and Mathys (**?**) or Foster and Stehrer (**?**) conduct a sectoral analysis of agglomeration economies for European countries. **?** focus, in particular, on manufacturing and financial services. Foster and Stehrer (**?**) find that agricultural sector has negative agglomeration economies. That is because density makes land price way more expensive and becomes a dispersion force for agricultural firms. Despite this particular

[7]Industry and sectors can be considered as synonymous in this paper.

[8]However, this proposition tends to confuse between concentration and agglomeration concepts. Agglomeration is linked to specific dense locations contrary to concentration index. A sector can be spatially concentrated but not located in agglomerated areas.

result, these two papers show that there exist positive agglomeration economies for all sectors, but their magnitudes differ according to sectoral features. Contrary to Brulhart and Mathys (**?**) or Foster and Stehrer (**?**), we exploit sectoral productivity data on a much finer spatial scale.

Other studies focus on sectoral heterogeneity of agglomeration economies. **?** introduces road congestion in agglomeration economies evaluation in the United Kingdom. He finds that some specific industries show concave relation between density and productivity: manufacturing, construction, hotel trade, transport & storage. In this sense, road congestion can offset 30% of agglomeration economies in the evaluation. Without road congestion, total agglomeration economies would be much higher. However, he also shows that for few sectors like banking or business service, agglomeration economies are constant or increasing with congestion.

Introducing interaction terms

There exist several ways to discuss how agglomeration economies may vary with other factors. While Brulhart and Mathys (**?**) or Foster and Stehrer (**?**) run several estimations for each sector, **?** or **?** insert an interaction term in their specification that necessarily includes the density variable.

? investigate how the magnitude of agglomeration economies on urban productivity varies according to local human capital stock. They find that not only human capital has a direct effect of productivity, but it also increases the positive effect of density on urban productivity. The interaction term is called the density of human capital. Their study is based on all 363 U.S. metropolitan areas data. The authors use the average output per worker between 2001 and 2005 to represent urban productivity. The agglomeration variable uses a population-based density. They show that the positive effect of human capital stock is higher in knowledge-intensive sectors as Service, Information or Finance. As in **?**, they introduce individual heterogeneity. The authors insert an interaction term to enlighten heterogeneity both among occupational status and sectors. They try to

understand who benefit most from agglomeration economies among workers. More precisely, **?** build two dummy variables: worker's occupational status and worker's sector. Their specification contains an interaction term with the product of these two dummies and density. In summary, workers with a high wage occupation benefit more from density than the others.

In this paper, the aim is to stress sectoral heterogeneity on pollution intensity. In this sense, we try to understand how pollution can affect agglomeration economies. The objective is close to **?** and **?**: we assess the magnitude of agglomeration economies according to a potential source of heterogeneity.

1.2.4 Air pollution in empirical studies

The impact of atmospheric pollution on location in urban context has been evaluated several times for households[9]. However, air pollution in the evaluation of agglomeration economies has been introduced only by Drut and Mahieux (**?**). This paper has been determinant in our work. They compare a standard estimation of agglomeration economies with an estimation integrating NO_X emissions. The difference corresponds to the net effect of agglomeration decreased by the pollution effect. As a first step, they find that employment density is negatively correlated with polluting emissions. Once they introduced pollution variable in the estimation, the impact of agglomeration economies decreases by more than 13% compared to standard estimates. The mechanisms behind are the following. Pollution damages labor productivity and higher density decreases pollution. Because they ignored the negative impact of density on pollution, previous studies over estimated the direct impact of density on productivity. Not only agglomeration increases productivity, but also decreases pollution. The latter is the indirect positive effect of density on productivity through a decrease of pollution.

Drut and Mahieux (**?**) do not introduce a cost of agglomeration, but on the contrary they insert a positive impact of density on productivity that is usually not taken into

[9]We have presented the strategy and the main results of this hedonic price literature in paper 2.

account (a decrease of pollution through an increase density has a positive impact on productivity). The negative correlation between density and pollution found by Drut and Mahieux (**?**) is essential in their work. However, the relation between density and pollution remains an open question in the literature, as we have seen in paper 2. In their case, density is computed at the level of employment areas, while they measure pollution at the level NUTS 2[10]. Yet, if a negative relation was found between density and pollution, we would get the reverse result: standard agglomeration economies measures are under-evaluated. The relation between density and pollution is not obvious and would deserve deeper analysis (Zeng and Eastin, **?**; He *et al.*, **?**. Because the authors want to build their analysis on health effect of pollution, they point out that an appropriate indicator could be the number of days when pollution exceeds certain limit.

Drut and Mahieux (**?**) focus on NO_X emissions from all sources, knowing that half of NO_X emissions comes from transports in France. Contrary to their study, our dataset covers all sectors but only emissions from highly polluting firms.

China constitutes a contemporary example of *arbitrage* between agglomeration with its benefits and environmental risks. In China, annual environmental cost has been evaluated to 8% of national GDP (**?**). Another striking fact concerns the number of most polluted towns in the world: 16 over first 20 are Chinese cities en 2001 (**?**). The interest can be seen through the number of recent publications on Chinese data. Nevertheless, French case remains also interesting. In France, plant locations of industries as metalworking or refineries and environmental awareness are older phenomena. These characteristics represent an advantage because we are looking for capitalized effects of density on productivity and of local polluting industrial emissions.

? find contrasted results about the link between density and pollution concentration or emissions. In their work, the relation corresponds to an inverse U-shape: pollution

[10]NUTS 2 corresponds to one level of *classification of territorial units for statistics* ;there are 27 NUTS 2 (French regions) over the dataset period they use.

emissions of SO_2 increase up to a certain level of density and decrease afterward. They also analyze SO_2 pollution intensity which corresponds to total annual emissions divided by gross industrial output. While density seems to be mainly positively linked to industrial pollution emissions, it is not the case with pollution intensity. This work makes us aware on the difficulties to get simple results on the links between density and pollution.

2 Data and variables

Our dataset covers all French employment areas between 2003 and 2013 in *metropolitan* France. In total, we consider at first: observations with 277 employment areas on 11 years for 79 sectors[11]. From the dataset DADS, we extract information on number of plants, average wage per sector and employment zones over the period. DADS dataset has the advantage to be exhaustive and reliable on the period. We also exploit IREP dataset in which yearly polluting emission information at the plant level is available. Plants belong to this dataset because they are considered as polluting plant and must declare every year their emissions. In the following section, we describe the most commonly used variables in the general framework -without pollution- to estimate the effect of employment density on productivity per worker. We also describe more precisely the *pollution* variable we build for the study.

The choice of spatial unit

A fine spatial scale seems to be determinant to investigate the effects of agglomeration and to identify the effect of polluting firms on agglomeration economies magnitude. We choose to measure productivity and all other variables at the level of *French employment areas*. While numerous studies are conducted at a larger scale with European

[11]However, we have in total only 180 763 observations. First, some sectors are not active in all employment zones. Moreover, the number of plants by sector in an employment zone may be very small. In this case, we could not have information on this sector in the employment zone scale because INSEE wants to make sure that we can not identify some specific plant information on wages.

regions on NUTS 3 level[12] (for example, Ciccone and Hall **?**), or on NUTS 2 level (see Brulhart and Mathys, **?**; Foster and Stehrer, **?**) ; we choose employment areas because agglomeration effects and perception of polluting emissions seem to decrease relatively quickly with distance.

French *employment areas* ("Zones d'Emploi" in French) are defined by INSEE, we choose to work with the latest version 2010 that fits well with our sample. Their borders are defined by economic mechanism as commuting patterns around a city. *Employment areas* are economically homogeneous and have lower border effects compared to administrative units. There are 304 employment areas in our dataset[13]. Using *employment areas*, in this kind of studies, has been initiated by **?** and adopted by the following works: Di Addario and Patacchini (**?**), **???**, **?**, among others. An interesting feature of our dataset is that all the variables can be calculated at the same spatial scale, including productivity, pollution and human capital variables[14].

Productivity and density

To assess the impact of agglomeration on productivity, we need some reliable information on the local level of both *agglomeration* and *productivity*. From an empirical point of view, the variable *agglomeration* is usually represented through the outcome, i.e. the *density*. However, we remind that the outcome and the process of agglomeration are different from a theoretical point of view. Because of a large heterogeneity of spatial extent of the geographical units, using a continuous variable such as *density* appears to be the best solution to reduce problems about mis-measurement. This measure has first been introduced by Ciccone and Hall **?**. While density of population has been used, we believe that employment density is more relevant in our study. It better corresponds to the level of local economic activity. Technically, we divide the number of jobs by the surface areas of employment zone.

[12]NUTS 3 levels corresponds to "*département*" and NUTS 2 level corresponds to "*région*" in France.

[13]We exclude DOM-TOM areas because of boundaries effects and the very different spatial context.

[14]Keeping the same spatial unit for measuring productivity and pollution is central. This is a key difference with other paper in the literature.

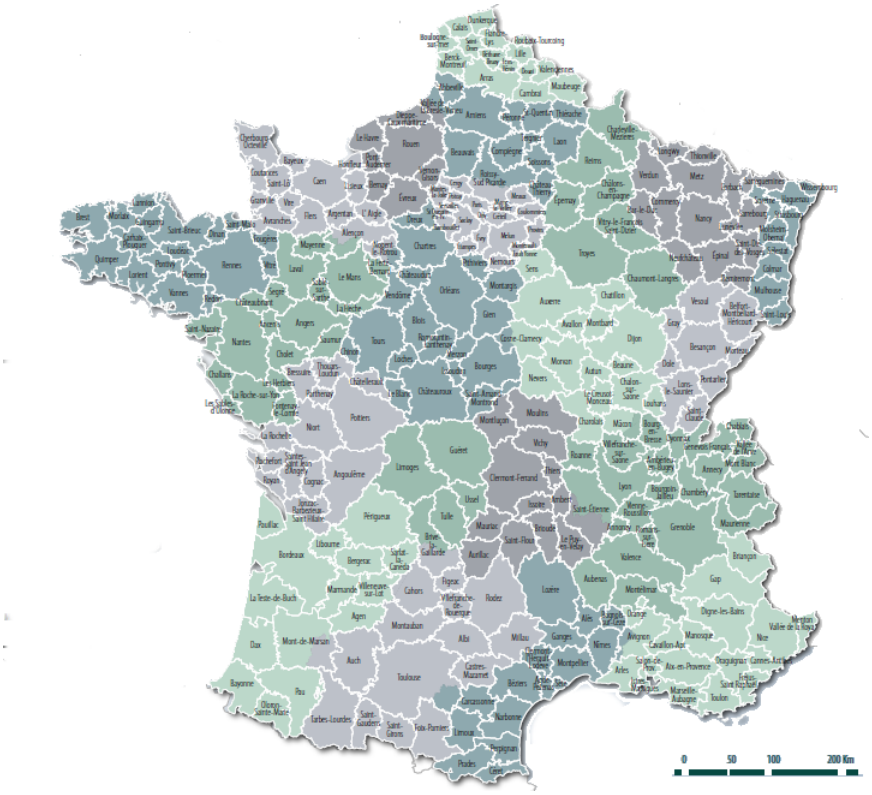


Figure 1. French employment areas, source : INSEE, Datar, 2011

Then, our measure of *productivity* is based on workplace-based data on nominal wages. We focus our analysis on labor productivity rather than *TFP* for several reasons. First, such information on wages is easily available; moreover, the correlation between these two measures of productivity is very high (**?**) and *TFP* has some notable drawbacks previously noted (section 1.2.2 page 10). To our knowledge, this paper, is the first empirical analysis matching polluting emission from Irep dataset and administrative data on wages (DADS). We match sectoral polluting emission, at the employment zone level, with sectoral wages at the same spatial scale. This matching makes possible the evaluation of the impact of pollution emissions on the amplitude of agglomeration economies for polluting sectors. Number of jobs and wages by employment areas are extracted from the DADS (Données Annuelles de Déclarations Sociales). The DADS is a declaration procedure which must be completed by all businesses that employ staff. In this document, which serves both fiscal and social administrative purposes, employers provide annually and for every establishment certain information pertaining to their

establishments and employees. We first have 86 sectors, but we keep only 79 sectors because locations of agriculture or extractive activities are too dependent on natural resources location. Most variables in our specification such as labor productivity information and quantity of polluting emissions are given at the industry levels.

Finally, to distinguish the agglomeration effect from the scale effect, we add a variable of spatial unit size: the *surface* of employment area from 2010 classification. Indeed, the size of the zone may influence directly productivity, because it represents the extent of local market and a certain ability to create interactions through proximity, with *density* considered as given. In this way, it is easier to understand and identify both density and size impacts[15].

Control variables

In order to get reliable coefficient of density, we introduce other local productivity determinants. *Market potential* of one area is the sum of "market" opportunities raised from all other neighboring zones. Indeed, if a plant wants to trade his production, it will consider all potential buyers around. We expect then a positive effect of market potential on productivity that depends on spatial extent considered. In theory, the home market effect is also a key factor of agglomeration (**?**); more precisely, the accessibility to a large market favors agglomeration of economic activities. The *market potential* represents also the fact that agglomeration economies may spill over. Therefore, it takes into account the spatial extent of agglomeration effects, because external markets can influence local productivity.

We define market potential as the sum of the density of the other areas, weighted by the inverse distance to these areas. We use euclidean distance between centroids of employment zones. $MP_{i,t}$ is the market potential of the employment area i at date t is

[15]However, to include both variables *surface* and *density* requires to be careful on the interpretations of estimated coefficients; see more details in appendix on this question, page 48.

calculated *à la* **?**:

$$MP_{i,t} = \sum_{j \neq i} \frac{dens_{j,t}}{dist_{i,j}}, \tag{2}$$

with $dens_{j,t}$ the density of employment zone j and $dist_{i,j}$ the distance between employment areas i and all the others j. As density of the zone, market potential from all other zones is likely to be endogenous.

The picture is not complete if we do not consider the industrial structure of the local economy. The heterogeneous effect of agglomeration economies can be partially captured by one variable: *diversity*. As *density*, *land area* or *market potential*, the variable *diversity* does not depend on a particular sector, but it is a local determinant of productivity. This indicator is inspired by Jacobs's intuition (**?**): that diversity of firms and sectors favors mutually interesting exchanges an a dense zone. The index of industrial *diversity* is constructed as the inverse of Herfindahl index.

$$Diversity_{i,t} = 1/H_{i,t} \ ,$$

with $H_{i,t} = \sum_s \frac{emp_{i,t,s}}{emp_{i,t}}$ the Herfindahl index, and $emp_{i,t,s}$ number of jobs of employment area i in sector s at time t.

We are conscious that Herfindahl index has some drawbacks. For instance, its values are influenced by the number of sub-categories, i.e. the number of industries (as we choose relatively detailed industry classification). Moreover, it doesn't take into account productive concentration by considering the number of plants and their size per industry. Yet, several arguments encourage us to use this index. First, it is the most commonly used in this literature (a few examples among others: Combes *et al.*, **?**; Martin *et al.***?**; Drut and Mahieux, **?**) and it allows to compare with benchmark studies. Thus, using an index like Ellison and Glaeser (**?**), which controls for productive concentration, would add a source of bias with the productivity variable. Indeed, industry level of productivity may influence plant size and productive structure in the

sector.

Education variables complete our set of control variables,with the aim to introduce human capital externalities. To build those variables, we use data from General Census of Population fully available on INSEE website at municipality level for population above 16 years old. We extract information on years: 1999, 2006 and 2011. Five categories of schooling achievements are established: *Without diploma*, *Middle school*, *High school* and *University*. To fit with the dataset 2003-2013, we identify levels of schooling for each employment zone and for each year in this way: 2003 to 1999 level, 2004-2008 to 2006 level, 2009-2013 to 2011 level.

Without considering such variables, we would overestimate the effect of density over productivity. Indeed, since education is one of the main local determinant of productivity, omitting this variable could lead to mis-identification of density effect[16]. However, we are aware that education might be endogenous: a high wage city can locally provide better education.

Pollution by sector

The study focuses on the years from 2003 to 2013 because data on polluting emissions are available on this period. IREP dataset is fully available on line[17] and contains yearly declarations of emissions from the main polluting plants in France.

Regional services of Environment Ministry audit polluting plants and assess their data reports. We keep only the 9 atmospheric polluting emissions among other polluting

[16]Stock of human capital is considered as exogenous in our work. We need individual data to go beyond this hypothesis. Sorting effect may play an important role in this case (**?**), that is why the use of individual level data -if available- can permit to control for this source of heterogeneity.

[17]On the website http://www.georisques.gouv.fr. This dataset has been very rarely used in empirical work. One exception is the work of **?** for equivalent US data in which he evaluates the market effects of the information availability on house market. The US program is called *Toxic Release Inventory*. In the French case, **?** employ IREP dataset to assess spatial and social inequalities toward environmental risks.

emissions and wastes reports[18]. We use the following list of IREP dataset pollutants: carbon monoxide (CO), carbon dioxide (CO_2), methane (CH_4), lead (Pb), mercury (Hg), cadmium (Cd), nitrogen oxide (NO_X), sulfur oxide (SO_X) and solvents and other volatile organic compounds ($COVNM$). These 9 pollutants are the most represented in terms of numbers of emitting plants in the IREP dataset. For every pollutant above, *Pollutant* variables count all the emissions from IREP plants in each sector and employment zone, emitted in one year.

Table 1 shows general classification of polluting emissions we chose to focus on. It displays three main categories established by **?**.

	Pollutants
Acidification and eutrofication	SO_2, NO_X, $COVNM$, CO
Greenhouse effect	CO_2, CH_4
Metal contamination	Cd, Pb, Hg

Table 1. Overview of polluting emissions and their consequences

Table 2 gives an overview of sectoral sources for polluting atmospheric emissions in our dataset. These features depend partially on plants that are present in IREP dataset. Two main sources are not included in our study: residential and individual transport emissions. However, one can observe that such information shares common trends with **?** data concerning emissions from firms, which is the French reference in terms of atmospheric pollution emissions.

To summarize, heavy metal contamination is mainly issued from manufacturing and chemical. Metalworking industry is the first sector for pollutants Pb and Cd, while Hg is mainly emitted in chemistry. Then, among pollutants that generate acidification and

[18]Our dataset is valuable for several reasons: first, agencies check the authenticity of emission reports. We also believe that information of this dataset is meaningful because these reports are partially used to levy tax called TGAP (see appendix in paper 2 for more details, page **??**) depending of the levels of pollution emissions for Sulfur oxide, Hydrochloric acid, Nitrogen oxide, Non-methane hydrocarbons, Solvents and other volatile organic compounds, fine particulates (PTS), polycyclic aromatic hydrocarbons.

	CH_4	CO	CO_2	$COVNM$	Cd	NO_X	Pb	SO_2
Manufacturing with coking	0,68	23,31	14,80	19,10	4,20	9,54	0,66	13,87
Chemical and metalworking	16,49	75,14	72,73	74,17	87,49	79,07	94,38	84,81
Electricity and water	74,02	-	9,17	3,10	6,67	8,09	2,39	0,90
Storage, trade and others	-	-	0,16	0,88	-	-	-	-
Services	0,47	-	2,19	2,76	-	2,81	0,11	0,13
Administration, public service	8,34	-	0,56	-	-	0,28	0,12	-
Creative activities	-	-	-	-	-	-	-	-

Table 2. Polluting emissions by sector of IREP plants in 2003 (in percentage)

eutrofication, the first emitters are the following: coking for SO_2 emissions, non metal mineral industry for NO_X, manufacturing industry for $COVNM$ and metalworking for CO. Finally, if CH_4 is mostly emitted by waste-processing industry (except farming, not included in our dataset), chemistry and mineral industry represent the most important emitters of CO_2. This information depends on our dataset where residential and transport are not concerned. As discussed in the introduction, industrial emission levels have decreased and the repartition among sectors has changed since 2003.

3 Empirical strategy

We test the following conditional hypothesis: do agglomeration economies decrease with the industrial polluting atmospheric emissions? To do so, we implement agglomeration economies evaluation with an interaction term and exploit a panel data set measuring local wage at the sectoral level between 2003 and 2013. For the main industrial pollutants, we use the sectoral quantity of pollutants emitted as an indicator of the degree of pollution. To identify this effect, we exploit two sources of variability: among employment areas-*spatial variability* and among years-*time variability*. Since we look at the causal impact of agglomeration on a local outcome such as the wage level, we face

endogeneity bias on our main explanatory variables. To tackle this issue, we implement an instrumental strategy which geological and instrumental variables.

3.1 Empirical specification

We regress sector average wage on employment density at French employment zone scale. We add several control variables such as: market access, index of sector composition and local skill composition. All variables are measured at industry and employment areas levels, calculated with log to get marginal effects and to smooth our data. Using variables in log also allows to compare our results with the majority of prior studies using log-log specification. The general specification is the following:

$$\log w_{ist} = \beta_0 + \beta_1 \log dens_{it} + \beta_{30} \log MP_{it} + \beta_{31} \log area_i + \beta_{32} \log div_{it} + \beta_{33} X_{it} + \gamma_s + \mu_t + \epsilon_{ist}, \tag{3}$$

where w_{ist} is the average wage in employment zone i, $dens_{it}$ represents the employment density in zone i, $area_i$ its surface, X_{it} corresponds to local human capital endowment, γ_s and μ_t are respectively an industry fixed effect and a time fixed effect, and ϵ_{ist} the error terms.

Table 3 displays the correlation between variables. As expected the variable *productivity* is positively correlated with local *density*, which is a necessary condition for positive agglomeration economies. All the other variables seems to have a positive correlation with *productivity*, except *area*, which is insignificant. We can also notice that *market potential* is highly correlated with *density* and it suggests a strong interconnection between both variables. As **?** predicted, the index of *diversity* is also positively correlated to *productivity*. We use two types of fixed effects. We consider year fixed effects to control for economic phenomena and temporal shocks that affect all sectors. For instance, the economic crisis of 2008 decreases production and therefore polluting emissions. Moreover, pollution emissions have also been reduced over this period be-

variables	$\log w$	$\log dens$	$\log MP$	$\log area$	$\log div$
$\log w$	1				
$\log dens$	0,265 ***	1			
$\log MP$	0,1642***	0,493***	1		
$\log area$	0,0025	-0,327 ***	- 0,399***	1	
$\log div$	0,1491 ***	0,389 ***	0,133***	0,307***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3. Correlation matrix

cause of an increase of general awareness toward environmental damages and the raise of regulations in Europe. Such temporal phenomena cannot disturb our estimation if we introduce year fixed effect[19].

Then, industry fixed effects are introduced. Industry fixed effects allow us to focus on pollution effect and to make our results independent from sectoral structures (**???**). For instance, waste processing industry and sector of R&D have very different production strategy and the magnitude of sectoral agglomeration economies would be impacted by industry structures. Yet, we need to isolate the effect of polluting emission levels from industry structure effect.

Finally, in order to identify the effect of density on local productivity, we analyze the variability of density across employment areas. With the same reasoning, to establish whether polluting emissions have an effect of agglomeration economies, we use "between employments areas variability" of polluting emissions within a sector.

Why we do not control for local fixed effects?

Even if it would permit to control all unobserved local determinants of productivity, we choose not to introduce local fixed effect for the same reasons. Even if it would permit to take into account all unobserved local determinants of productivity (such as historical amenities), local fixed effects have serious limits in this context. First, if the specification includes local fixed effect, the estimation of density effect is built on

[19]We need to suppose that the crisis affected all sectors in a similar way. This assumption is frequently made in such studies (for instance **?**).

time-related variation of density. As we explained it before, variability of density is likely to be tracked by measurement errors. Second, inserting local fixed effect is not sufficient for tackling the endogeneity issue. Indeed, it only address the endogeneity from all local unobserved features that were omitted. Local fixed effects do not help tackling reverse causality.

3.2 Specification with polluting emissions

We want to fit as much as possible to the results of NEG models with polluting firms. The aim is to explore whether pollution emissions affects economics agglomerations. Do benefits for an industry to locate in a dense areas decrease with its amount of pollution emitted? Contrary to Drut and Mahieux (**?**), our purpose is not to assess the impact of the local level of pollution on local productivity[20], but rather to evaluate the extent to which agglomeration economies are lower for polluting sectors.

To figure out whether generating pollution decrease the benefit of agglomeration for polluting firms, we look at the effect of the variable *pollution* on agglomeration economies. We introduce the variable *pollution* in an interaction term with *density*. For each pollutant, the specification is the following:

$$\log w_{is} = \beta_0 + \beta_1 \log dens_i + \beta_2 \log dens_i \times \log pollution_{is} + \beta_3 \log pollution_{is} + \beta_4 X_i + \gamma_s + \mu_t + \epsilon_{ist}, \quad (4)$$

where $pollution_{is}$ represents the level of pollution emitted by sector s in employment zone i, and x_i represent all control variables at the employment zone scale. In this section, we are interested in the coefficients β_1 and β_2, because they measure agglomeration economies accounting for polluting emissions effects. To sum up, we assess how the relationship between agglomerations and the sectoral productivity varies with the

[20]In paper 2, we presented literature trying to assess pollution impact on productivity and empirical challenges to identify pollution effects on worker productivity.

amount of pollution emitted by the sector.

3.3 Dealing with the endogeneity of local density

In the literature, two types of endogeneity concerns are specified: local and individual levels of endogeneity. First, at the local level, two sources of potential endogeneity threaten our identification strategy: missing variables and reverse causality. There might be local factors influencing both density and productivity that are difficult to capture. An area with a high average productivity attracts firms and workers, which increases density. GMM estimator with panel dataset could also be a way to correct for unobserved heterogeneity, however, it is not pertinent in our case[21]. Second, endogeneity issues at the individual levels will not be treated in this paper because of data availability issues. Yet, there might be for instance some self-selection due to unobserved abilities.

In a NEG model, productivity and agglomeration reinforce each other. So it would not be realistic to consider the level of agglomeration as exogenous in this empirical study. Endogeneity issues bias the estimate of the coefficient for three of our main variables: density, market potential and interaction term. We need to choose the best way to tackle this endogeneity bias. According to our dataset and in line with previous literature, we conduct an empirical strategy based on instrumental variables. We instrument endogenous variables with historical values of density (Ciconne and

[21]Despite our panel dataset, we do not use panel dimension for several reasons. First, exogeneity of GMM instruments appears to be very problematic in our case. GMM approach uses lagged values of the variables of interest as instruments. The variable $Density_{i,t}$ in each area i at date t is clearly serially correlated. One solution could be to increase the lag, but it does not fit to our dataset. Indeed, GMM approach implies to restrict already to 8 over 10 years time period, because the two first years would be used for instrumentation. The restriction will be higher if we increase the lag. Second, as we have a relative short period of time, time variability of employment area density is low and not meaningful.

Moreover, we do not use GMM because this method captures only short-run variations, and we prefer to focus on long-term variations. From an economic point of view, short-run variations are not our purpose in this work because we want to identify capitalized agglomeration economies in wage - variations that might be influenced by the presence of polluting firms. To sum up, although panel data is an usual strategy toward endogeneity issue, it is not an appropriate method in this paper due to the time span of our dataset and due to our problematic.

Hall, **?**) and geological variables (Rosenthal and Strange, **?**). An appropriate instrument should be correlated with instrumented variable *density*, but not correlated with the error terms, i.e. not correlated with missing local determinants and not explained by *productivity*.

Historical instruments

We use population densities from 1831 and 1861 for each current employment area[22] to instrument employment *density* and the same values divided by the inter-zone distances to instrument *market potential*. Historical data on density or population are highly linked to current values of density. Indeed, some factory plant locations or housing of 19th remain until today, and they create at least a certain inertia in terms of density (**?**). To be an accurate instrument, historical values should not be correlated with current local productivity. It is likely to be the case if we consider some historical instruments with a long lag. Because economy structure has changed deeply and several shocks as war have occurred since 1831, labor productivity should not be correlated with these historical instruments[23]. With table 4, we can see that log *density 1831* and log *density 1861* are good predictors of current log *employment density*.

	log *employment density*
log Density 1831	0.196
log Density 1861	0.236

Nb of observations: 277

Table 4. R^2 of univariate regressions: Historical values

Geological instruments

[22]These data are extracted from **?** with her permission.

[23]We are conscious that there exist a growing trend criticizing historical instruments of density. For instance, according to **?**, institutions, some natural endowments or good climate are very long lasting. These might represent omitted variables influencing both productivity and instruments. For the same reason, cultural and historical amenities may create a phenomenon of sorting. These instrument seem to offset only partially this issue. Historical censuses were conducted for any municipality with population above 5000 inhabitants. We drop employment zones for which historical values are not available, 277 employment zones remain.

Geological variables come from the European Soil Database (ESDB) and include soil features, altitude, water stocks and many other characteristics. Those variables had a great impact on past population location because agriculture activities strongly contributed to the emergence of cities. Then, geographical variables on soil do not directly affect the labor demand or the current productivity in most sectors. One exception might be farming but this sector is excluded from the dataset. Moreover, agriculture weight in national economy decreased sharply and the land fertility does not drive local wealth anymore. Combes and Gobillon (**?**) point out that geological instruments do not have as strong explanatory power as historical instrument. However, the over identification test is relevant if we have instruments of different nature and not only past population densities as instruments. Moreover, our results are very unlikely to be impacted by a violation of the exclusion restrictions since we test for the stability of our coefficient by switching among different instruments.

All geological variables, cited below in table 5, are composed of several dummies. We regress log *employment density* against the 6 sets of variables. Then, we choose to present only the R^2 for each regression to compare the prediction power, in comparison with historical variables. In general, we can notice that geological instruments do not predict as well as historical instruments log *employment density*. Yet, three of them seem to explain relatively well employment density (see tables 5): Depth to rock, Dominant parent material and Subsoil water capacity.

	log *employment density*
Depth to rock (3 dummies)	0.131
Subsoil water capacity (6 dummies)	0.139
Soil differentiation (4 dummies)	0.126
Soil erodibility (5 dummies)	0.129
Topsoil mineralogy (6 dummies)	0.127
Dominant parent material (14 dummies)	0.132

Nb of observations: 277

Table 5. R^2 of univariate regressions: Geological values

Table 6 displays the first stage of the 2SLS estimation: it the OLS regressions of the endogenous variable on our instruments. We do not report all coefficients for every dummies; however, it must be noted that at least one dummy for every set of instruments is significant at 5%. Partial R^2 are reported in columns (1) to (8). We develop a weak instrument tests (Stock and Yogo, **?**) with the F-test reported in the table. Because results of F-test are clearly above the critical threshold suggested by Stock and Yogo (**?**), historical values of population can be considered as strong instruments of employment density. Geological instruments have F-test results that confirm the fact they are weaker instruments (**?**). At the end, we will only keep Depth to rock and Soil erodibility, which have a F-test above the threshold established by Stock and Yogo (**?**).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log Density 1831	0.019***							
	(0.002)							
log Density 1861		0.020***						
		(0.002)						
Depth to rock	-	-	✓	-	-	-	-	-
Subsoil water capacity	-	-	-	✓	-	-	-	-
Soil differentiation	-	-	-	-	✓	-	-	-
Soil erodibility	-	-	-	-	-	✓	-	-
Topsoil mineralogy	-	-	-	-	-	-	✓	-
Parent material	-	-	-	-	-	-	-	✓
R squared	0.196	0.236	0.131	0.139	0.126	0.129	0.127	0.132
F-test	67.23	84.94	13.72	8.774	13.07	8.013	7.819	3.338

Note: Dependent variable: log (employment density); Standard errors in parentheses, and $* = p < 0.10$, $** = p < 0.05$, and $*** = p < 0.01$; nb of observations: 277

Table 6. First Stage estimates of Density

The variable *market potential* and the interaction term -in the second specification with pollution (see equation (4))- are plagued by the same endogeneity issue: we need to instrument them as well. In total, we will treat this identification issue on three potentially endogenous variables.

The instrument of *market potential* will use spatial lags of historical values (see equation (2), page 21). Moreover, we need to define a relevant instrument of the

interaction term. **?** investigate how education achievement influences urban economies and face the same problem. To construct a relevant instrument, they use the instrument of density multiplied by education term to build the instrument of interaction term. We will proceed in the same way to instrument our interaction term : : instrument of density multiplied b pollution.

4 Results

We present the aggregate results and pollutant specific results in the following sections. The first section aims to present the results of the basic framework of agglomeration economies evaluation. Then, we introduce our variable of interest with an interaction term to check whether agglomeration economies vary with polluting emissions.

4.1 Aggregated effects

General results are displayed in table 7 that is composed of 7 columns, one by regression. The method followed in the first three columns assumes that *density* variable is uncorrelated with the error term. However, as we explained in section 3.3, several bias can occur. To tackle this problem, we insert two types of instrument, both at the employment zone spatial scale: historical population density and geological variables. For the geological variables, we keep the local indexes of depth to rock and soil differentiation because of the outcomes of F-test cited above. We check that our estimation strategy is consistent to a change of instruments. Indeed, when we introduce each instrument one by one, the coefficient of density remains similar. It attests that the results are not too local and that the exclusion restrictions is very unlikely to be violated. Because the instruments based on historical values show stronger explanatory power, we keep them for the rest of the study.

Table 7 shows that employment density clearly favors labor productivity, the coefficient is significant at 1% level for all our specification. Following density variable, all

control variables have a positive influence on productivity. Differences in the level of agglomeration are high among employment zones in France. We find that the inter-decile ratio $\frac{P90}{P10}$ is around 10 in 2003. For instance, following specification (1) , predicted agglomeration economies are 14% higher for the employment zone of Mulhouse (P90) than for the employment zone of Auch (P10)[24].

Beside non negligible effect on density coefficient, industry fixed effects increase sharply the R^2 (from column (1) to column (2)). Such result also confirms the interest to insert both year and industry fixed effects in the estimation. To introduce local human capital makes employment density effects smaller as suggested by **?**.

The coefficient of *density* decreases from 0.055 (see columns (2)) to 0.043 (see columns (4)) once we take into account endogeneity through historical instruments, in accordance with the literature. We also account endogeneity bias that can affect the variable *market potential*. Our instrumentation strategy affects a little less its coefficient with a decrease from 0.041 to 0.038, considering the regressions without education.

To sum up, we choose to interpret the results of column (7) because endogeneity issue on density has been treated and human capital is included as a control variable. We think regression (7) captures well the treated mechanisms and constitutes the best regression at this stage. We can read the coefficients of column (7) in the following way. In the regression of aggregate agglomeration economies with instrumental variables, an increase of 1% in density implies an increase of 0.022 % in local labor productivity. In other words, predicted agglomeration economies are 4 % higher for the employment zone of Mulhouse (P90) than for the employment zone of Auch (P10). Concerning the control variables, human capital variables have the expected positive effect on labor productivity. We take the value "*Without Diploma*" as the reference. For instance, an increase of individuals with a high school or a university degree favors local productivity, compared to those without any diploma[25]. *Market potential* is positive and highly

[24]We follow **?** and apply the following formula: $\left(\frac{P90}{P10}\right)^{\beta_1} - 1) \times 100$.

[25]We remind that the magnitude of each control variable coefficients is not our aim ; yet, we will focus on heterogeneous pollution effect of agglomeration economies.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			$\log w_{is}$				
	OLS			2SLS			
log Density	0.063***	0.055***	0.029***	0.043***	0.062***	0.066***	0.022***
	(0.003)	(0.003)	(0.003)	(0.007)	(0.014)	(0.010)	(0.005)
log Market potential	0.042***	0.041***	0.051***	0.038***	0.050*	0.039**	0.047***
	(0.008)	(0.008)	(0.005)	(0.011)	(0.029)	(0.020)	(0.007)
log Area	0.037***	0.029***	0.012***	0.020***	0.035***	0.035***	0.006*
	(0.003)	(0.003)	(0.003)	(0.004)	(0.007)	(0.007)	(0.003)
log Diversity	0.019*	0.014	0.021**	0.045***	-0.004	-0.010	0.032***
	(0.011)	(0.011)	(0.010)	(0.014)	(0.027)	(0.022)	(0.011)
Middle school			-0.658***				-0.649***
			(0.185)				(0.189)
High school			0.466**				0.417**
			(0.198)				(0.211)
University			0.515***				0.612***
			(0.072)				(0.076)
Historical instrument	–	–	–	✓	–	–	✓
Soil differentiation	–	–	–	–	✓	–	–
Depth to rock	–	–	–	–	–	✓	✓
Year fixed effect	✓	✓	✓	✓	✓	✓	✓
Industry fixed effect	–	✓	✓	✓	✓	✓	✓
Observations	187,856	187,856	187,856	187,856	187,856	187,856	187,856
R-squared	0.181	0.635	0.643	0.633	0.634	0.634	0.642

This table reports the OLS (1-3) and 2SLS (2-5) regressions coefficients of employment density over sectoral wages. The local level of density is instrumented by local historical urban population in 1831 and 1861 and geological variables described at the employment zone scale, market potential is instrumented by spatial lags of urban population in 1831 and 1861. The sample is restricted to employment areas with non-null urban population in 1831 (277 employment zones). All regressions include a constant. Standard errors are in brackets and they are clustered at the employment area level; $* = p < 0.10$, $** = p < 0.05$, and $*** = p < 0.01$.

Table 7. Aggregate Agglomeration Effects

significant: 10% of increase in *market potential* implies 0.46 % in *productivity*. The surface of the zone does not have a very significant coefficient, mostly because of the variable definition (see appendix, page 48): an increase of 10% in surface induces a 0.06 % increase of productivity. In the literature, the variable can have some contradictory results. Our study tends to confirm **?**'s intuition and we find that a 10% of increase in diversity implies 0.32 % increase in productivity.

4.2 Agglomeration and polluting emissions

4.2.1 Interaction between agglomeration and pollution

In section 1, we see that pollution affects the amplitude of agglomeration economies. To assess empirically this prediction, we estimate the impact of pollution emissions on agglomeration economies ; we remind the rationale behind this specification in the following equation:

$$\log w_{is} = \beta_0 + \beta_1 \log dens_i + \beta_2 \log dens_i \times \log pollution_{is} + \beta_3 \log pollution_{is} + \beta_4 X_i + \gamma_s + \mu_t + \epsilon_{ist},$$

For each sector s in location i, we introduce an interaction term between the total quantities emitted by sector s in location i and the local density in location i.

Table 8 displays the results of our interaction specification for 9 pollutants. As explained in detail by **?**, coefficients in interaction models should not be interpreted independently. To interpret the interaction terms, we need to consider that the condition *Pollution* modifies the effect of density on productivity and calculate proper marginal effects. In table 8, we see that 4 over 9 pollutants have an interaction term negative and significant. To be able to interpret the results, one needs to focus on graphs presented in figure 2.

Figure 2 illustrates how the marginal effects of agglomeration on sectoral productiv-

ity change with the sectoral level of pollution. Any particular point on the black lines corresponds to:

$$\frac{\partial w_{is}}{\partial Density_i} = \beta_1 + \beta_3 \log Pollution_{is},$$

with β_1 and β_3 corresponding coefficients of the empirical specification equation (4) (page 27). The dotted line depicts the 5 % confidence interval. Figure 2 is very informative on the magnitude of agglomeration economies. For non-polluting sectors economic agglomeration is high and highly significant. For instance, figure 3b shows that agglomeration economies are around 0.02%, meaning that an increase of 10 % in the level of local density lead to an increase of labor productivity around 0.02 % for non-polluting sectors). However, sectoral agglomeration economies shift towards zero as sectoral pollution increases. For instance, figure 3b shows that, if a sector emits more than 70 grams of lead per day, then his agglomeration economies are around zero. For very high level of sectoral emission, agglomeration economies turn out to be negative. For instance, for high levels of *lead* pollution, figure 3b wages decrease with agglomeration (negative agglomeration effects significantly different from zero). The intuition is the following: given the risks of lead pollution, less dense places clearly represent a better alternative for productivity for those sectors using lead. We find similar results for others source of pollution such as: CO_2, SO_2, and NO_2.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					$\log w_{is}$				
					2SLS				
	CH_4	C0	$C0_2$	COVNM	Cd	Hg	NO_2	Pb	SO_2
log Density	0.020***	0.020***	0.019***	0.019***	0.020***	0.020***	0.020***	0.020***	0.020***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
log Pollution	0.002	0.000	-0.001*	0.000	-0.000	-0.008*	-0.001***	-0.004***	-0.001
	(0.001)	(0.001)	(0.000)	(0.000)	(0.002)	(0.004)	(0.001)	(0.002)	(0.001)
log Pollution× log Density	0.005***	0.009***	0.008***	0.006***	0.015***	0.026***	0.009***	0.013***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.001)	(0.002)	(0.001)
log Market potential	0.048***	0.048***	0.048***	0.048***	0.048***	0.048***	0.048***	0.048***	0.048***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
log Area	0.005	0.005	0.004	0.005	0.005	0.005	0.005	0.005	0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
log Diversity	0.031***	0.032***	0.031***	0.031***	0.031***	0.031***	0.031***	0.031***	0.031***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Middle school	-0.710***	-0.708***	-0.675***	-0.693***	-0.713***	-0.709***	-0.695***	-0.708***	-0.702***
	(0.198)	(0.199)	(0.198)	(0.197)	(0.199)	(0.199)	(0.198)	(0.199)	(0.198)
High school	0.391*	0.389*	0.395*	0.402*	0.390*	0.390*	0.386*	0.389*	0.389*
	(0.219)	(0.220)	(0.220)	(0.218)	(0.219)	(0.220)	(0.220)	(0.220)	(0.219)
University	0.623***	0.623***	0.631***	0.627***	0.622***	0.623***	0.627***	0.624***	0.626***
	(0.075)	(0.076)	(0.076)	(0.075)	(0.075)	(0.075)	(0.076)	(0.076)	(0.075)
Year fixed effect	✓	✓	✓	✓	✓	✓	✓	✓	✓
Industry fixed effect	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	187856	187856	187856	187856	187856	187856	187856	187856	187856
R-squared	0.638	0.638	0.640	0.639	0.638	0.638	0.639	0.638	0.639

This table reports the 2SLS (1-9) regressions coefficients of employment density over sectoral wages. The local level of density is instrumented by local historical urban population in 1831 and 1861, market potential is instrumented by spatial lags of urban population in 1831 and 1861. The sample is restricted to employment areas with non-null urban population in 1831. All regressions include a constant. Standard errors are in brackets and they are clustered at the employment area level; $* = p < 0.10$, $** = p < 0.05$, and $*** = p < 0.01$.

Table 8. Interaction between Sectoral Pollution and Agglomeration

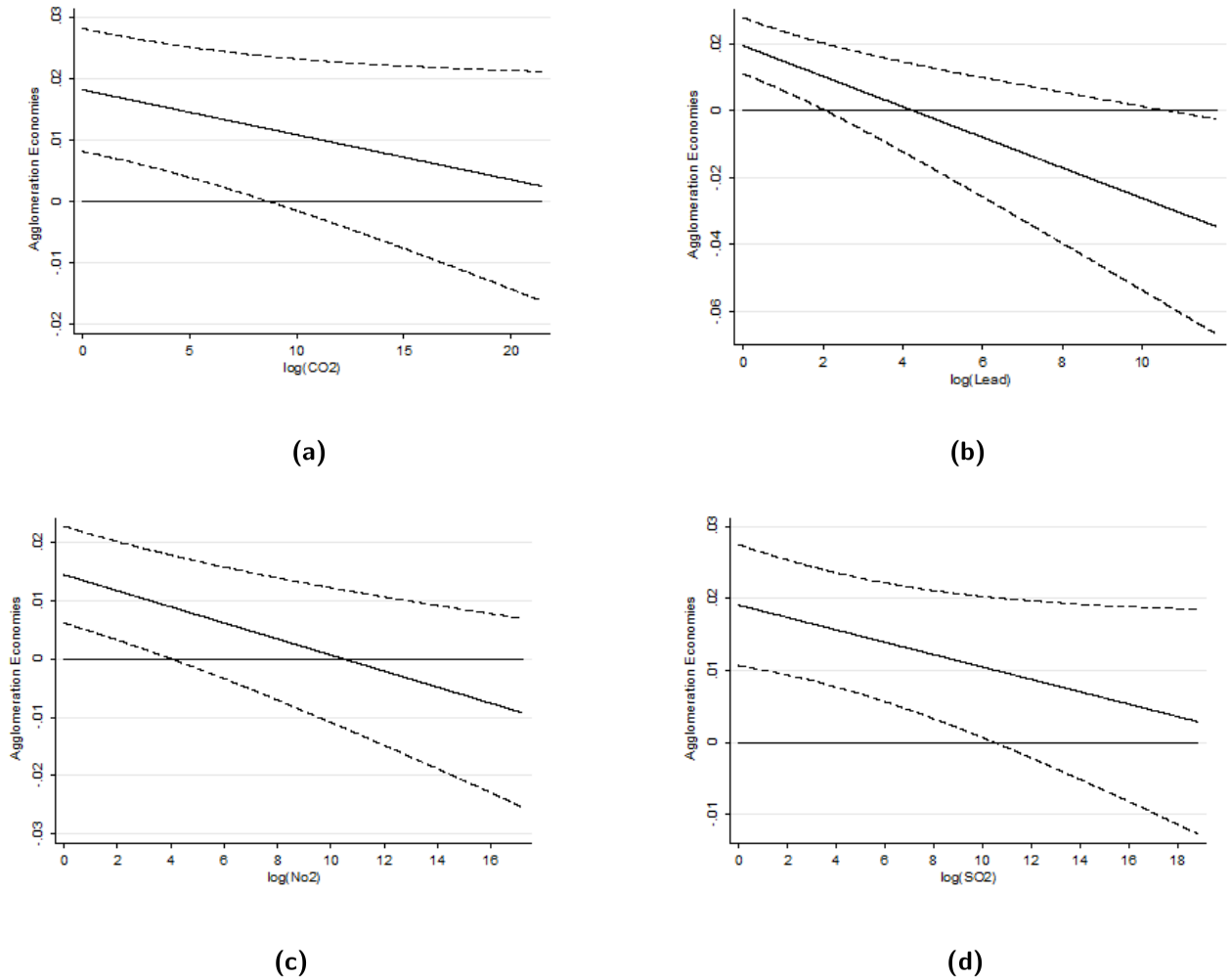


Figure 2. Interaction model: Impact of pollution emissions on economics agglomerations

These plots display the effect of sectoral level of emission on economics agglomeration. They illustrate how the marginal effect of density on wages changes with the sectoral level of emission. Any particular point on the black lines corresponds to: $\frac{\partial w_{is}}{\partial Density_i} = \beta_1 + \beta_3 \log Pollution_{is}$). The doted line depicts the 10 % confidence interval.

To sum up, as sector polluting emissions increase, firms belonging to this sector will benefit less from agglomeration. However, our results might be biased due to the endogeneity of pollution. In following subsection, we discuss this potential bias and propose a strategy to deal with this issue.

4.2.2 Dealing with the endogeneity of pollution

Our measure of the local sectoral degree of pollution (namely variable $pollution_{is}$) is potentially endogenous. There may exist some omitted variables that influence both local labor productivity and local polluting emissions. For instance, a local negative economic shock can affect both local wages in a specific zone, and pollution emissions via a decrease in the final demand for goods.

A suitable strategy to address the endogeneity issue is to use variations in pollution emission that is a plausibly exogenous to the evolution of the local sectoral context. We instrument the local level of pollution in a specific sector by a *shift-share* instruments[26]. With *shift-share* methodology, we consider that pollution *shift* is constituted of several components (or *share*, hence the name of the methodology). Change of our *Shift-share* instrument contains variations that are uncorrelated with possibly endogeneous local-industry component. Our instrumental variable approach focuses on national year-to-year changes in pollution emission. This strategy is based on the following idea: national variations in polluting emissions are uncorrelated with some specific local economic shocks. Yet, such variations are more related to technical progress, e.g. "green" technical progress. National changes in pollution emissions are translated into expected levels of pollutant emissions by sector and employment zones. We use the levels of local sectoral emissions in 2003 to construct the predicted local levels of

[26]Shift-share instruments are popular in labor economics. Since **?**, local employment growth rate is instrumented by its prediction: an interaction between local industry employment shares with national industry employment growth rates.

emissions. The formula is the following:

$$\widehat{Pollution}_{ist} = Pollution_{is,2003} \times (1 + \Delta Pollution_{fr,2003-t}) \ ,$$

where $\widehat{Pollution}_{ist}$ is the predicted quantity of pollutants emitted by sector s in the employment area i and year t. $\Delta Pollution_{fr,2003-t}$ are the national rates of emission growth between 2003 and year t. $Pollution_{is,2003}$ is the level of emitted pollutant in location i in the sector s for the year 2003. In this dataset, 2003 constitutes the first year for which such geographic detail is available.

This prediction is independent from local specific sectoral shocks. Two identification assumptions are made. First, we assume that the initial pollution in 2003 is not influenced by some omitted variables affecting local productivity in the future. The second identifying assumption is the following: national rates of emission growth are not influenced by the quantity emitted by a particular sector in a particular spatial zone. Such assumption is easily fulfilled, since no local sector is sufficiently large to track national rates of emissions. We construct such predicted pollution for all of our nine pollutants.

Table 9 shows the first stage of the 2SLS estimation. In the first stage, actual levels of emission of sector s in area i are regressed on the predicted levels of emission and the other control variables in table 8. All instruments are highly significants. Partial F-test critical values for the excluded instruments vary between 240 to 35 depending on the pollutants. All instruments are above critical values tabulates by **?**; therefore, we can predict sectoral levels of pollutants by using the national rates of emission growth.

Table 10 presents the basic results using instrumental variables. Pollution is instrumented by the fitted values obtained from the regressions in table 9. Results are broadly similar to those in the specification in which pollutant emissions are not instrumented (see table 8). The interaction term of CO_2, lead and NO_2 are still significant and keep the same magnitude. However, once SO_2 emission is instrumented, its interaction term

and degree of agglomeration are no longer significant. Coefficients after instrumentation of Hg (mercure) and Cd (cadmium) -heavy metal contamination- are now significant. Generally, the stability of our coefficients across specifications demonstrate that our findings are not tracked by endogeneity issues.

	(1)	(2)	(3)	(4)	(5)	(7)	(8)	(9)	(10)
	$CH4_{ist}$	CO_{ist}	$CO2_{ist}$	COVNM_{ist}	Cd_{ist}	Hg_{ist}	$NO2_{ist}$	Pb_{ist}	$SO2_{ist}$
log Predicted Pollution	0.598***	0.459***	0.681***	0.551***	0.427***	0.509***	0.683***	0.453***	0.788***
	(0.039)	(0.082)	(0.019)	(0.021)	(0.073)	(0.073)	(0.030)	(0.041)	(0.044)
Control variables	✓	✓	✓	✓	✓	✓	✓		✓
Industry fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓		✓
F-Test	234.9	31.52	1326	677.2	34.46	48.02	517.3	121.4	323.6
Observations	171054	171055	171056	171057	171058	171060	171061	171062	171063
R-squared	0.335	0.186	0.499	0.355	0.056	0.242	0.474	0.185	0.427

The table shows the regressions of emission of pollutants for sector s located in area i for year t on the predictions (instrumental variables).
Control variables are variables present in table 8, namely local level of density, Market Potential, Area, Diversity, Middle School, High School and University.
In 2003 (ie. reference year in our instrumental strategy), for some area and some sector, there are only one polluting firms by observation. We delete those observations. They do not constitute a credible baseline for our specification, because they are too dependent on the situation of a specific firms. Thus, the number of observations is not constant between pollutants.
All regressions include a constant. Standard errors are in brackets and they are clustered and the employment area level; $* = p < 0.10$, $** = p < 0.05$, and $*** = p < 0.01$.

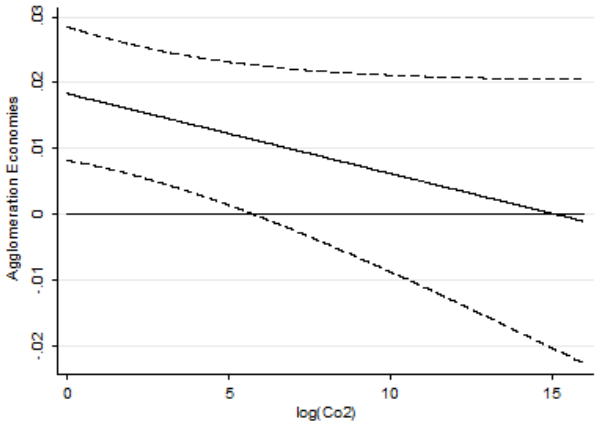
Table 9. First Stage Estimates of Local Sectoral Level of Pollution

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
					$\log w_{is}$ 2SLS				
	CH_4	CO	CO_2	COVNM	Cd	Hg	NO_2	Pb	SO_2
log Density	0.020***	0.020***	0.018***	0.018***	0.020***	0.020***	0.020***	0.020***	0.019***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
log Pollution	0.010	0.014**	0.010***	0.011***	0.067***	0.034***	0.010***	0.022***	0.012***
	(0.007)	(0.006)	(0.001)	(0.001)	(0.014)	(0.011)	(0.002)	(0.008)	(0.002)
log Pollution× log Density	0.003	0.001	-0.001*	0.001	-0.030***	-0.033**	-0.002*	-0.017**	-0.000
	(0.006)	(0.018)	(0.001)	(0.001)	(0.011)	(0.016)	(0.001)	(0.007)	(0.001)
log Market potential	0.048***	0.048***	0.048***	0.048***	0.048***	0.048***	0.048***	0.048***	0.048***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
log Area	0.005	0.005	0.004	0.004	0.005	0.005	0.005	0.005	0.005
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
log Diversity	0.031***	0.032***	0.031***	0.031***	0.032***	0.031***	0.031***	0.031***	0.032***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Middle school	-0.705***	-0.705***	-0.661***	-0.681***	-0.701***	-0.700***	-0.690***	-0.695***	-0.698***
	(0.198)	(0.199)	(0.199)	(0.195)	(0.200)	(0.199)	(0.199)	(0.200)	(0.197)
High school	0.390*	0.389*	0.392*	0.413*	0.373*	0.385*	0.383*	0.378*	0.389*
	(0.219)	(0.219)	(0.223)	(0.214)	(0.223)	(0.222)	(0.221)	(0.223)	(0.217)
University	0.625***	0.623***	0.634***	0.630***	0.629***	0.624***	0.628***	0.627***	0.628***
	(0.075)	(0.076)	(0.076)	(0.074)	(0.076)	(0.076)	(0.076)	(0.076)	(0.075)
Industry fixed effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Year fixed effects	✓	✓	✓	✓	✓	✓	✓		✓
Observations	171,054	171,055	171,056	171,057	171,058	171,059	171,060	171,061	171,062
R-squared	0.637	0.638	0.640	0.638	0.631	0.637	0.639	0.637	0.638

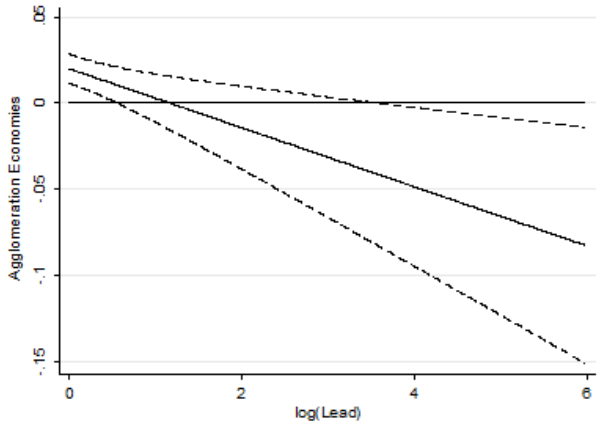
This table reports the 2SLS (1-9) regressions coefficients of employment density over sectoral wages. The local level of *density* is instrumented by local historical urban population in 1831 and 1861, *market potential* is instrumented by spatial lags of urban population in 1831 and 1861. The level of pollution is instrumented by its predicted value. The sample is restricted to employment areas with non-null urban population in 1831. All regressions include a constant. Standard errors are in brackets and they are clustered at the employment area level; $* = p < 0.10$, $** = p < 0.05$, and $*** = p < 0.01$.

In 2003 (i.e. reference year in our instrumental strategy), for some areas and some sectors, there are only one polluting firms per observation. We need to delete those observations. They do not constitute a credible baseline for our specification, because they are too dependent on the situation of a specific firm. Thus, the number of observations is not constant across pollutants.

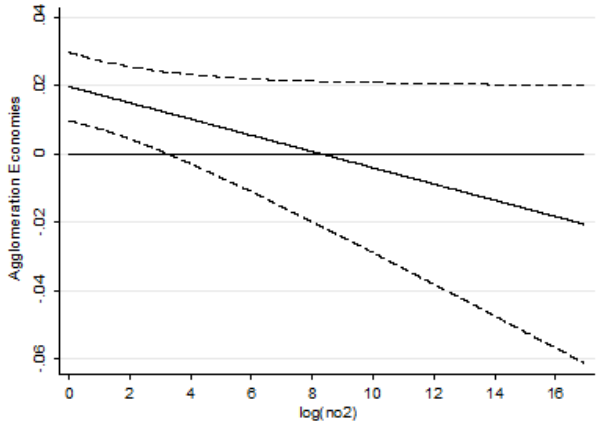
Table 10. Interaction between Pollution and Agglomeration when both Density and Pollution are Instrumented



(a)

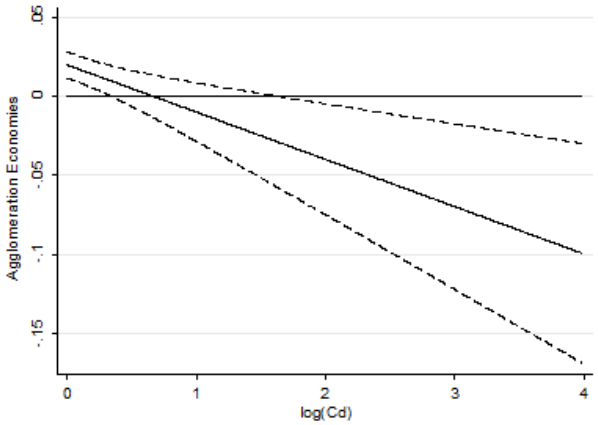


(b)

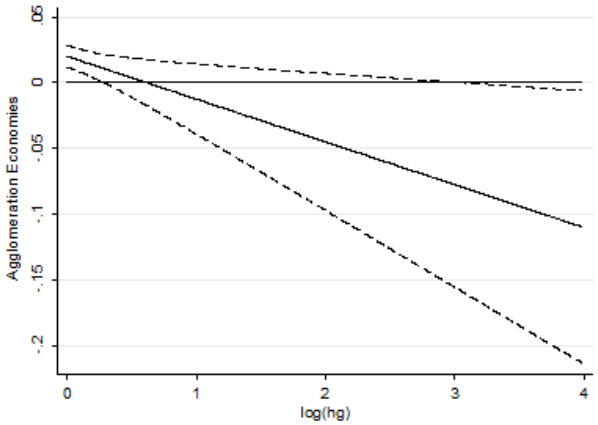


(c)

Figure 3. Interaction model: Impact of pollution emissions on economics agglomerations when both agglomerations and pollution are instrumented


Agglomeration Economies
.05
0
-.05
-.1
-.15
0
1
2
3
4
log(Cd)

(d)


Agglomeration Economies
.05
0
-.05
-.1
-.15
-.2
0
1
2
3
4
log(hg)

(e)

Conclusion

The paper provides an estimation of agglomeration economies once polluting emissions are introduced. To do so, we exploit French labor productivity and pollutant emissions between 2003 and 2013. We use both DADS and IREP datasets to get information on wages and emissions at fine industry and spatial scales.

We find that firms have lower interest to locate in a dense area as their polluting emissions increase. More precisely, increases of emissions of NO_2, lead, CO_2 or SO_2 interfere in the way firms benefit from agglomeration economies. From a certain threshold-defined for each pollutant, wages start to decrease with employment density.

Three main variables might be endogenous: employment *density*, *market potential* and *pollution*. In line with the literature, we instrument the first two with historical values of population density. For the variable *pollution*, we use a method adapted from **?**. It corrects for endogeneity bias due to omitted variables influencing both local productivity and local level of pollutant emissions.

Our study brings new insights on negative aspects of agglomeration on labor productivity. While in most studies, only positive impact of agglomeration is enlightened, we concentrate on polluting emissions as source of heterogeneous effect of agglomeration.

Several limits and orientations for future research can be underlined. Despite the fact that we chose to focus on agglomeration economies evaluation, the next step will be to tackle the question of firm location choice. Then, in this work, the aim was not to focus on channels affecting agglomeration economies. To better understand how pollution can interfere with agglomeration economies, one way would be to focus on health effect and to deal with endogeneity issue of pollution exposure toward productivity.

References

Appendix paper 4

Discussion about some coefficients

Few precisions need to be done about coefficients for some control variables. Because we are considering several times the same variables in our ln specification, their coefficients should not be directly interpreted. Here is the explanation for *area* and *specialisation* variables.

Specification equation can be synthesized as:

$$\ln prod_{zs} = \alpha + \beta \ln dens_z + \upsilon \ln area_z + \theta \ln spe_{zs}$$

Knowing that:

$$\beta \ln dens_z + \upsilon \ln area_z + \theta \ln spe_{zs} = \beta \ln emp_z - \beta \ln area_z + \upsilon \ln area_z + \theta \ln emp_{zs} - \theta \ln emp_z$$

We can then conclude by:

$$\ln prod_{zs} = \alpha + \boldsymbol{\rho} \ln emp_z + \boldsymbol{\gamma} \ln area_z + \theta \ln emp_{zs}$$

with $\boldsymbol{\rho} = \beta - \theta$ and $\boldsymbol{\gamma} = \upsilon - \beta$.

The last line pictures what we observe after running our regression. So if ρ is negative for instance it doesn't mean that density has a negative effect, but more that both can be positive but θ larger than β.

With the same approach, if γ is negative or non significant, it would be wrong to conclude that *area* does have a negative impact on productivity. **?** summarize int this way: "When using density and land area, agglomeration gains exist when any of the estimated coefficients is significantly positive".