Local labour markets and spatial determinants of overeducation

Florian Fouquet

Lemna
EA 4272

Document de travail du LEMNA N° 2023-02

Avril 2023

Local labour markets and spatial determinants of overeducation

Florian Fouquet

Local labour markets and spatial determinants of overeducation

Florian Fouquet*

February 2023

Abstract

This paper investigates the spatial dimension of overeducation and its incidence on local labour markets. Spatial Durbin panel models, using both static and dynamic specifications, are run on French employment areas between 2009 and 2017. I find that the lack of job opportunities affects overeducation both locally in the neighbouring areas. I also find that the geography of local labour markets does not have the same impact on educational mismatch for all workers, with young workers being less affected by the situation in other areas. Finally, I find that overeducation is related to opportunities in rural areas, and to competition in urban areas.

Keywords: overeducation, educational mismatch, local labour markets, spatial heterogeneity, spatial panel models

JEL Classification: C33, J24, R23

*LEMNA EA4272 (Laboratoire d'Economie et de Management de Nantes-Atlantique), Nantes University, France; Address: IAE Nantes - Economie & Management Chemin de la Censive du Tertre BP 52231 44322 Nantes Cedex 3 FRANCE; Mail: florian.fouquet@univ-nantes.fr

I would like to thank participants in the 19th International Workshop on Spatial Econometrics and Statistics, the 17th annual conference of the research federation CNRS TEPP, the 38th *Journées de Microéconomie Appliquée* and the LEMNA seminar for their comments on this paper.

1 Introduction

Since the 1980s, most advanced countries have seen a strong increase of overeducation, defined by McGuinness (2006) as the situation where "an individual possesses a level of education in excess of that which is required for their particular job". Kucel (2011) showed that overeducation almost doubled in the Netherlands, the UK and the USA between the 1980s and the beginning of the 2000s. Overeducation involves an underutilisation of resources that is costly both for the workers and for the firms, entailing lower wages (Verdugo & Verdugo, 1989), less job satisfaction and lower productivity (Tsang et al., 1991), as well as higher turnover rates (Sloane et al., 1999). This has led to a large literature exploring both its causes and its consequences.

One plausible explanation for overeducation is the large increase in educational attainment (Groot & Maassen van den Brink, 2000). Although less than 20% of 25-64 year-olds had a tertiary degree at the end of the 1990s in Europe[1], this had risen to 37.5% by 2020. Moreover, for 25-34 year-olds, this share rose from about 21% in 1999 to 45% in 2020. However, recent studies challenge the importance of this driver in explaining overeducation (Croce & Ghignoni, 2012; Davia et al., 2017). These studies show that an increased supply of skilled labour does not directly lead to the emergence of overeducation. On the contrary, factors preventing demand from adjusting play a major role, in particular institutional factors and the economic climate. To better understand the mechanisms at stake, it is therefore crucial to investigate other determinants of overeducation. The aim of this paper is to document its spatial dimension and the influence of local labour markets.

A few studies have investigated the role of geographical mobility and the regional labour market context at the individual level (Büchel & Van Ham, 2003; Hensen et al., 2009; Jauhiainen, 2011). In particular, they show that individual mobility reduces the risk of overeducation, and that local labour market characteristics, such as unemployment rate or the structure of employment, are linked to educational mismatch. While the general labour economics framework involves individuals searching for jobs on a global market, most workers actually have limited spatial flexibility, and search for jobs around their residential locations, on local labour markets. Simpson (1992) explains that, if a job seeker is on a local labour market with no suitable jobs available, she or he has three options. The first is not to participate in the market, *i.e.* to stay unemployed. The second is to accept a job for which she or he is overeducated. The third is to geographically broaden her or his job search.

[1]More precisely, in the European Union's OECD 23 members. Source: OECD (2022), Population with tertiary education (indicator). doi: 10.1787/0b8f90e9-en (Accessed on 9 August 2022)

The relevance of geography, local contexts and spatial interactions in explaining overeducation remains understudied in the literature. In one of the few studies exploring these dimensions, Tselios (2013) showed that the local economic context can explain differences in the risk of overeducation between regions. Rodríguez-Pose and Tselios (2011) found that educational inequalities and attainment are related to characteristics of the region, but also of the neighbouring regions. Moreover, the distribution of skills demand and supply is not random (Murphy et al., 1991; Rodríguez-Pose & Tselios, 2011). Indeed, regions compete to attract highly educated workers and industries (Rodríguez-Pose & Vilalta-Bufí, 2005), because of the importance of these workers in economic growth (Martin & Ottaviano, 2001; Murphy et al., 1991). This reinforces the necessity to take into account the local context and neighbouring labour markets when studying overeducation.

My investigation of the spatial dimension of overeducation and the effects of local characteristics of the labour market has a threefold objective. Firstly, to determine whether or not the local labour market context influences overeducation. Several studies have explored the effects of labour market characteristics, such as unemployment (Büchel & Van Ham, 2003; Croce & Ghignoni, 2012; Guironnet, 2006), economic dynamism (Croce & Ghignoni, 2012; Tselios, 2013) or the distribution of firms (Ghignoni & Verashchagina, 2014; Nauze-Fichet & Tomasini, 2002). However, most of these studies addressed these effects by considering large areas. I contribute to this literature by addressing the smaller scale of local labour markets, rather than regional or national ones.

Secondly, I seek to assess the extent to which a local labour market's situation influences overeducation on neighbouring markets. Indeed, Manning and Petrongolo (2017) showed that local shocks impact labour markets at a much broader scale than their own area. Furthermore, Patacchini and Zenou (2007) showed that there is strong spatial dependence in local unemployment rates. To my knowledge, however, no similar investigations have been performed with regard to overeducation rates. Yet, according to Simpson (1992), overeducation and unemployment are alternative situations, meaning that a similar mechanism can be expected to lie behind both these outcomes. Moreover, Langella and Manning (2022) showed that individuals "flee" less dynamic labour markets (with higher unemployment), suggesting once again that the situation on one labour market impacts the situation on others.

Thirdly, I explore whether these geographical factors have the same impacts on overeducation for different gender and age subgroups of workers. The seminal paper of Frank (1978) regarding the spatial dimension of overeducation explains that women face a higher risk of

overeducation because of their limited spatial mobility. According to this "theory of differential overqualification", in a couple, it is the husband who looks for a job on the global labour market, whereas the wife looks for a job only on the local labour market of the area where her partner has obtained work. However, while empirical studies confirm that workers with higher spatial flexibility are less likely to be overeducated (Büchel & Van Ham, 2003; Hensen et al., 2009; Jauhiainen, 2011), there is no evidence that the spatial dimension leaves women more affected by educational mismatch than men (Büchel & Van Ham, 2003). Concerning young workers, they are known to be more overeducated that older workers (Baert et al., 2013; Meroni & Vera-Toscano, 2017). However, they are also more mobile. Moreover, Marinescu and Rathelot (2018) showed that although job seekers have a "distaste for distance", those more educated, who are on average younger, are more likely to apply for jobs located farther from their residence. My aim is thus to determine the mechanisms by which the local labour market context affects overeducation for these subgroups.

I pursue these three aims by analysing overeducation at a macro-level and estimating a panel model in which variables are observed at the French employment area level from 2009 to 2017. Spatial econometrics models are used to take spatial dependence into account. Three main results emerge from the analysis. Firstly, an area's labour market characteristics have a strong impact on overeducation. In particular, firms' demographic dynamics play a major role in explaining educational mismatch at the local labour market level. Secondly, there is clear evidence of the spatial dependence of labour markets. The results confirm the influence of neighbouring areas on educational mismatch. In particular, overeducation rates are higher when individuals face a lack of job opportunities on neighbouring labour markets. Moreover, these results suggest migratory behaviour by workers to avoid overducation: rates are about 0.3 points higher on average for each additional point of overeducation in the neighbouring areas in the previous period. Thirdly, not every demographic group is affected in the same way. Women appear to be more affected by competition for skilled jobs, while the geography of local labour markets appears to play less of a role in young workers' overeducation.

Section 2 describes the data and the variables, and provides some descriptive statistics. The model is detailed in Section 3. Section 4 provides and comments on the estimation results. Section 5 concludes.

2 Data and variables

2.1 Data

While many studies analyse overeducation at the individual level (Baert et al., 2013; Büchel & Van Ham, 2003; Hensen et al., 2009; Jauhiainen, 2011; Meroni & Vera-Toscano, 2017), I choose to explore it at an aggregate level in order to investigate the role of local characteristics and neighbouring local labour markets. Mobilising data from different sources, I rebuild a large set of variables regarding the French employment areas, defined by the French National Statistical Institute (Insee) as areas where most workers both live and work, and where firms can find most of their workforce. These areas are defined from commuting flows between municipalities, each municipality being included in only one employment area and indivisible, even when large. Employment areas are composed of neighbouring municipalities; while their population cannot be less than 5000, it should not be too large, ensuring their relevance to the study of local labour markets (see Appendix D for a cartography of French employment areas).

My main source is the *Professional activity* database of the population census conducted by Insee. This dataset provides individual information on education, labour market situation and demographic characteristics for around 7.5 millions workers for each year of the panel, by French employment area. I use these data to calculate the overeducation rates in each employment area, together with some local labour market characteristics such as employment by business sector, the share of individuals working in the area but living in another and indicators of skilled labour supply and demand. I also use the municipality-level database of the population census to retrieve information on local unemployment, as the *Professional activity* database only contains employed individuals. The municipality-based numbers of unemployed and active individuals are aggregated at the employment area level to calculate local unemployment rates.

The French census survey is not conducted annually for the whole population. One fifth of the sample is re-surveyed each year. Thus, ideally there should be a gap of five years between observations of a given area in the panel. Unfortunately, this is impossible because data are available only for years 2008 to 2017. I therefore choose to observe a given area every four years, which ensures that 80% of the census sample has been re-surveyed and avoids too much data persistence. I build my panel of 297 employment areas of metropolitan France[2] using years 2009, 2013 and 2017. Explanatory variables are time-lagged by one year to avoid a simultaneity problem (see Section 3) and are computed respectively for years 2008, 2012 and 2016.

[2]I exclude the seven employment areas of Corsica because of its insular nature.

Finally, I use the Directory of Enterprises and Institutions data (REE), which provides municipality-based information on the demography of private firms. From it, I calculate the number of microfirms (with less than 10 workers) and the number of new firms established in years 2008, 2012 and 2016 in each employment area.

2.2 Variables

2.2.1 Overeducation rates

Local overeducation rates are calculated using the population census, which indicates whether workers have a tertiary degree or not. An individual is here considered as overeducated if she or he has a higher education degree but works in an occupation not requiring one. Overeducation rates are then calculated as the ratio between the number of overeducated workers and the number of employed individuals in the area. This imperfect measure leads to underestimating overeducation: for example, individuals with an ISCED[3] 7 degree (long tertiary degree) but working in an occupation requiring ISCED 6 (short tertiary degree) are not identified as overeducated. This could bias results, if adding these unobserved overeducated workers changed the geography of overeducation. However, more detailed information about education is available for year 2017, allowing me to measure overeducation more precisely. I compute these overeducation rates and draw a map to check the robustness of this measure. While overeducation is now slightly higher, its spatial distribution does not substantially vary using this alternative measure (see Appendix A).

To determine whether an occupation requires a higher education degree, I rely on the *Professional families* classification of the French Ministry of Labour (Dares)[4]. For each of the 87 professional families[5], information about required qualifications is available from the Operational Directory of Occupations and Jobs (ROME) of the French state employment agency *Pôle Emploi*. There are nine levels of qualification in ROME, the three highest (Technicians, Supervisors, Managers and engineers) can be considered to require a tertiary degree. For some of the professional families, classification is straightforward because only these three levels are involved. For families with a wider range of required qualifications, I consider that a higher education degree is required if more than half of the workers in the occupation have one.

[3]International Standard Classification of Education

[4]The *Professional families* classification is built by linking socio-professional categories (PCS) and qualifications required as defined in the Operational Directory of Occupations and Jobs (ROME). *Professional families* are particularly appropriate to study employment in terms of qualifications, opportunities, mobilities and supply-demand relationships.

[5]The final total is 85 families, after excluding Armed forces and Religious occupations.

2.2.2 Explanatory variables

The choice of explanatory variables is driven by the theoretical background of overeducation. Human Capital Theory (Becker, 1964) stresses the importance of worker characteristics and institutional factors[6]. The Job Competition Model (Thurow, 1975) underlines the importance of job characteristics in explaining educational mismatches. Assignment models (Sattinger, 1993) can be seen as an intermediate framework between Human Capital Theory and the Job Competition Model. These models predict that job allocation is the result of an assignment problem and depends on both the distribution of workers and the distribution of jobs. McGuinness (2006) indicated that this framework is entirely consistent with overeducation, and I therefore include information on both the distribution of workers and the distribution of firms.

Included in the distribution of workers is the share of 25-34 year-olds in the active population, drawn from the population census. Overeducation is known to particularly affect young workers entering the labour market (Baert et al., 2013; Meroni & Vera-Toscano, 2017), so that at the aggregate level, overeducation may be higher in areas whose workforce includes more young workers. To take into account the spatial flexibility of workers (Büchel & Van Ham, 2003; Hensen et al., 2009; Jauhiainen, 2011), I include the share of individuals living outside the employment area of their workplace.

For the distribution of firms, I include the log of the number of microfirms (with less than 10 employees), since the probability of being overeducated has been found to decrease with size of firm (Ghignoni & Verashchagina, 2014; Guironnet, 2006). One possible explanation is that small firms are less likely to provide on-the-job training, in which case hiring an overeducated worker reduces the risk of choosing a candidate who lacks the required skills. Moreover, recruitment processes are more formalised in larger firms, reducing the risk of mismatch. Thus, overeducation is likely to be more of a problem in areas where firms are smaller. In addition, I include the log of the number of newly established firms as an indicator of the local economic dynamism of the areas. Since overeducation affects some sectors more than others (Nauze-Fichet & Tomasini, 2002), I also include the distribution of employment among the different sectors.

Finally, I include variables addressing the local labour market. To control for the effect of increased educational attainment (Groot & Maassen van den Brink, 2000), I use an indicator of

[6]In Human Capital Theory, there should not be any underutilisation of human capital. Thus, overeducation should not exist at the equilibrium. However, McGuinness (2006) indicated that, as a short-run phenomenon, overeducation remains consistent with this framework, in particular because of institutional factors that for instance prevent firms from adjusting their production processes to the new characteristics of the labour supply.

the imbalance between supply and demand for tertiary-level graduates in the active population. This variable is the ratio between the share of tertiary graduates in the labour force (proxy for the skilled labour supply) and the share of individuals in ISCO 1, 2 and 3 occupations in the labour force (proxy for the skilled labour demand). If this ratio is greater than one, there are too many graduates for the skilled jobs available; if this ratio is lower than one, not all jobs can be filled by the area's skilled labour force. In addition, I control for the share of temporary employment (Croce & Ghignoni, 2012; Ghignoni & Verashchagina, 2014). Indeed, according to the theory of turnover (Jovanovic, 1979), mismatches are a consequence of imperfect information. Repeated job seeking is therefore a way to improve matches, by obtaining more and better information. Thus, short-term contracts should help reduce overeducation. Furthermore, the unemployment rate is included for two reasons linked to the sign of the effect. A positive link between overeducation and unemployment could indicate an unhealthy local labour market, where matches are both rare and inefficient (Guironnet, 2006). On the contrary, a negative link could be a macro-level sign of the microeconomic mechanism of overeducation being a way to avoid unemployment (Ghignoni & Verashchagina, 2014).

2.3 Descriptive statistics

Table 1 presents the main characteristics of the sample in terms of overeducation rates for the years 2009, 2013 and 2017 (respectively 2008, 2012 and 2016 for the explanatory variables). Overeducation rates increased over the period, from 8.3% in 2009 to 11.5% in 2017 on average. The rise was higher at the top (+5 points for the maximum value) than at the bottom of the distribution (+3.4 points for the minimum value). As expected, there were large geographical disparities (Figures 1 and 2): overeducation was higher in the west of France and in the south, in particular the south-east, while it was relatively less marked in the center-north. Overeducation rates were higher for women than for men, with a larger increase over the period (9.9% in 2009 and 13.2% in 2017 for women, against 6.9% in 2009 and 9.2% in 2017 for men). The 25-34 year-olds constituted about a fifth of the labour force on average. As expected, overeducation rates for this subgroup were significantly higher than for the overall population, both at the beginning and the end of the period. However, the increase was slower (+1.9 points against +3.2 for the overall sample). The share of workers living outside the employment area of their workplace slightly increased between 2008 and 2016, possibly indicating that workers were broadening their job search area to find a better job.

Table 1: Sample characteristics

	2009 (2008)		**2013** (2012)		**2017** (2016)	
Variables	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Dependent variables						
Overeducation rate	0.083	0.016	0.099	0.017	0.115	0.018
Female overeducation rate	0.099	0.019	0.114	0.020	0.132	0.022
Male overeducation rate	0.069	0.014	0.079	0.015	0.092	0.016
Youth overeducation rate	0.143	0.025	0.152	0.025	0.162	0.026
Distribution of workers						
% of 25-34 year-olds in the labour force	0.214	0.019	0.207	0.018	0.200	0.018
% of workers living outside their area of work	0.182	0.097	0.192	0.098	0.204	0.101
Distribution of firms						
Number of microfirms per 100000 inhabitants	13282	4771	15575	5441	19279	6789
Number of newly established firms per 100000 inhabitants	1216	463	1912	661	2120	787
% of employment in the agricultural sector	0.054	0.038	0.051	0.036	0.049	0.035
% of employment in the industrial sector	0.175	0.070	0.161	0.064	0.155	0.063
% of employment in the construction sector	0.076	0.013	0.076	0.013	0.070	0.012
% of employment in the sales sector	0.215	0.035	0.215	0.035	0.214	0.035
% of employment in the services sector	0.169	0.042	0.175	0.043	0.179	0.044
% of employment in the education, health and public sector	0.311	0.046	0.322	0.046	0.333	0.048
Local characteristics						
Excess supply of tertiary graduates	0.879	0.059	0.929	0.059	0.992	0.062
% of temporary contracts	0.160	0.025	0.160	0.025	0.166	0.025
Unemployment rate	0.109	0.025	0.126	0.027	0.136	0.028
Observations	297		297		297	

Note : Raw values are reported for the number of microfirms and newly established firms rather than the log.
Overeducation rates are reported for years 2009, 2013 and 2017. The explanatory variables are lagged by one year to avoid simultaneity, thus, values are reported for years 2008, 2012 and 2016 respectively.
Source : French population census and REE

The distribution of employment among the different sectors varied slightly between 2008 and 2016, the share of the services and public sectors increasing at the expense of the agricultural and industrial sectors. Overall, services, sales, education, health and the public sector represent about 70% of employment on average. There was also a significant rise in the number of newly established firms, likely reflecting a more favourable economic climate at the end of the period than at the beginning. However, the mean unemployment rate increased by 2 points at the same time. The share of temporary contracts appears stable over the period.

On average, the indicator of imbalance between the skilled labour supply and demand was lower than one, pointing to more jobs requiring a tertiary degree than suitably qualified workers. However, this ratio increased sharply over the period, reaching almost one at the end of the period. Moreover, this average hides large disparities between employment areas, with the ratio ranging from about 0.7 to more than 1.15. This may indicate a spatial mismatch between jobs and workers, highlighting once more the importance of investigating the geographical dimension of overeducation.

3 Model specification

To investigate the local determinants of overeducation, I exploit the panel structure of the data to take into account unobservable heterogeneity across areas. Moreover, this increases the size of my sample by repeating observations over time. I start from the following model:

$$Y_t = X_t\beta + \alpha + u_t$$

where the dependent variable Y_t is the overeducation rate by French employment area in year t and X_t is the vector including local characteristics. The explanatory variables are time-lagged by one year to deal with a potential simultaneity problem[7]. As there are only two periods (2013 and 2017) in the panel[8], I run a poolability test to verify the inclusion of area effects. The null hypothesis is rejected (see Table 4), meaning that panel effects need to be taken into account. A Hausman test is run to determine whether to use random or fixed effects. The null hypothesis is rejected (see the last lines in Table 4), therefore I estimate a fixed effects model.

[7] Alternatively, I could address endogeneity through an instrumental variables approach. However, as I will be estimating dynamic spatial models, Maximum Likelihood estimators will be needed. Unfortunately, these estimators cannot deal with IV in the dynamic spatial models used here (Elhorst, 2003).

[8] Year 2009 is used in the DSDM model. To ensure comparability of results between the static and dynamic specifications, the static models are run on years 2013 and 2017 only.

3.1 Spatial specification

The spatial dimension of the data allows me to verify that the local labour markets are interconnected, *i.e.* to explore whether or not the situation in one area influences the situation in the neighbouring areas, and how. Although Manning and Petrongolo (2017) showed that local labour markets are not precisely defined geographical divisions, but overlapping areas, the structure of the data makes it impossible to take this directly into account. However, this structure allows for the use of spatial models to take into account interactions between neighbouring areas. Moreover, autocorrelation between neighbouring observations violates a necessary assumption of standard models and leads to biased estimators if this spatial structure is ignored (Cliff & Ord, 1973, 1981). There is spatial autocorrelation when neighbouring observations are not independent. Autocorrelation is positive (negative) if very similar (very different) values of the same variable are observed in neighbouring areas.

Table 2: Spatial autocorrelation of overeducation rates (Moran's I)

	2009		**2017**	
Spatial weight matrix	Moran's I	p-value	Moran's I	p-value
First-order contiguity	0.426***	0.002	0.383***	0.002
Second-order contiguity	0.340***	0.002	0.317***	0.002
3 nearest neighbours (3NN)	0.468***	0.002	0.435***	0.002
4 nearest neighbours (4NN)	0.452***	0.002	0.411***	0.002
5 nearest neighbours (5NN)	0.443***	0.002	0.408***	0.002
6 nearest neighbours (6NN)	0.426***	0.002	0.407***	0.002
7 nearest neighbours (7NN)	0.421***	0.002	0.403***	0.002
Inverse distance (max 100km)	0.380***	0.002	0.357***	0.003
Number of observations	297		297	

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

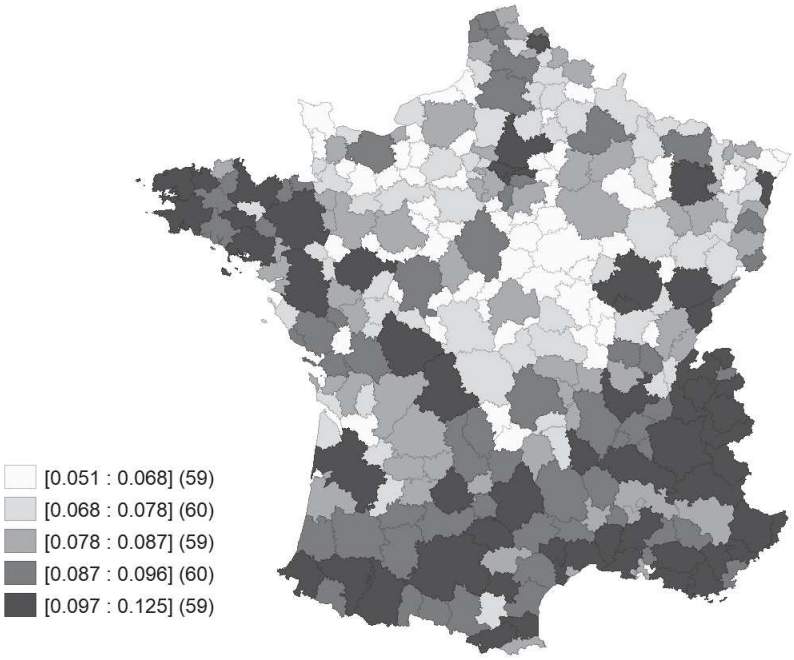
Figures 1 and 2 display overeducation rates by employment areas for years 2009 and 2017. Several clusters of areas show similar overeducation rates, particularly striking in the south-east of France, where almost all areas are coloured darker grey both in 2009 and 2017. To confirm the presence of spatial autocorrelation, I then calculate Moran's I coefficients (Moran, 1950)[9], not computed on the whole panel, but at a particular point in time using cross-sectional information. Table 2 presents the values of the Moran's I for years 2009 and 2017. These are always very significant, regardless of the spatial weight matrix.

[9]Moran's I is defined as follows:

$$I_W = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}; i \neq j$$

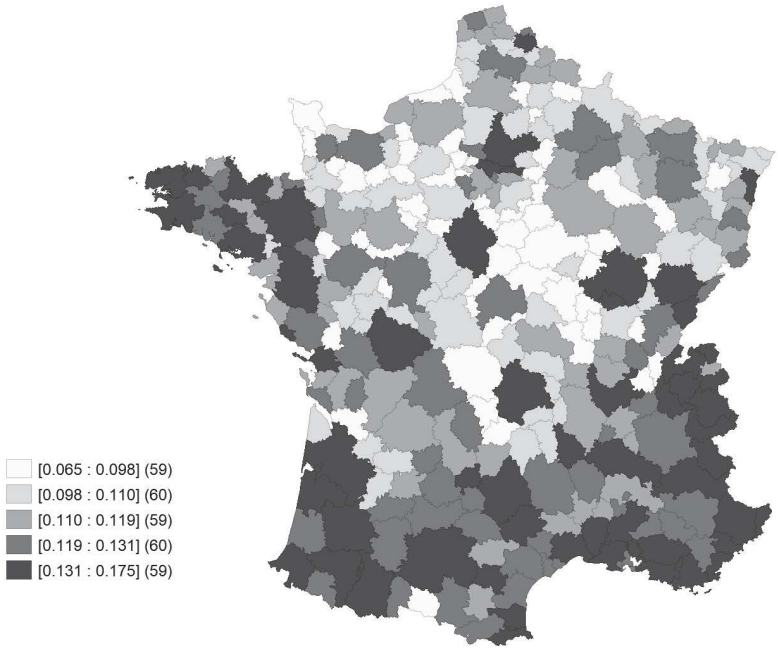
where n is the number of observations, y_i and y_j are the values of the variable in area i and j respectively and $\bar{y}$ the mean value of that same variable, and w_{ij} the element of the spatial weight matrix that defines the proximity between area i and area j.

Figure 1: Overeducation rates by French employment area - 2009



Note: Areas in white have the lowest overeducation rates, areas in dark grey have the highest overeducation rates.

Figure 2: Overeducation rates by French employment area - 2017



Note: Areas in white have the lowest overeducation rates, areas in dark grey have the highest overeducation rates.

I therefore turn to spatial econometrics models. While consideration of spatial dependence in panel models is a fairly recent development, a variety of models are designed to take spatial dependence into account (Elhorst, 2003, 2014). I follow the specification procedure proposed by Debarsy and Ertur (2010), based on Lagrange multiplier tests (LM-tests)[10], in order to

[10]These LM-tests are score tests based on the gradient of the likelihood function. When testing the SAR model

identify the presence of spatial autocorrelation on the dependent variable or in the error term. As a supplement, likelihood-ratio tests (LR-tests)[11] are performed to determine whether there is spatial dependence in the explanatory variables (Bouayad Agha et al., 2018). Results of the specification tests are reported in Table 3, for various spatial weight matrix designs.

They point to spatial autocorrelation on the dependent variable (SAR model, test 1 in Table 3) and in the error term (SEM model, test 2). The Spatial Durbin Model (SDM), which incorporates spatial lags of the independent variables and is a general form that includes these two models, is preferred to the SAR model (test 3). Given the results of the tests, I estimate a SDM, which can be written as follows:

$$Y_t = \rho W Y_t + X_t \beta + W X_t \theta + \alpha + u_t$$

where W is the spatial weight matrix. In the Spatial Durbin Model, coefficients cannot be interpreted directly. Such a model makes it possible to distinguish between direct effects, *i.e.* the average impact of a variable observed in a given area on overeducation in the same area, and indirect effects, *i.e.* the average impact on overeducation in one area of the explanatory variables observed in other areas (LeSage & Pace, 2009). The total effect, which is the sum of the previous two effects, is the average impact of a regressor on the dependent variable (see Appendix B for the calculation of these different effects).

3.2 Dynamic specification

Another concern is the persistence of overeducation. Since some individuals remain overeducated for a long time (Baert et al., 2013), overeducation at a given point in time is likely to be related to its past values, especially if new overeducated workers join those already mismatched. Therefore, I add a time lag of the dependent variable. As the model already includes spatial lags of the dependent variable and the regressors, a Dynamic Spatial Durbin Model (DSDM) is required. I test two versions of the DSDM model, the first including both a time lag and a space-time lag of Y (full DSDM model), and the second only a time lag of Y (denoted DSDM1 in the following). Both DSDM models are preferred to the static SDM (tests 4 and 5). Fi-

(respectively the SEM model), the null hypothesis is that the spatial coefficient ρ (λ) is equal to zero. When the null hypothesis is rejected, the spatial structure of the data needs to be taken into account.

[11] LR-tests are based on comparing the log-likelihoods of a constraint model, without a spatial term (or equivalently with the spatial coefficient equal to zero) and a non-constraint model where the coefficient is not equal to zero. Practically, these tests verify that the log-likelihood of the non-constraint model (in this case, including the spatial lags of the independent variables) is significantly higher than the log-likelihood of the constraint model. LR-tests are also used here to verify the inclusion of the time-lagged dependent variables.

nally, DSDM1 is tested against the full DSDM model. The results of test 6 for the inclusion of a space-time-lagged dependent variable show that the full DSDM is preferred to DSDM1. I therefore estimate the full DSDM model, which can be written as follows:

$$Y_t = \tau Y_{t-1} + \rho W Y_t + \eta W Y_{t-1} + X_t \beta + W X_t \theta + \alpha + u_t$$

Like the SDM, DSDM models give direct and indirect effects, but they also distinguish between short-term and long-term effects. The calculation of these different effects is detailed in Appendix B.

The choice of spatial matrix relies on an approach based on log-likelihood function values (Stakhovych & Bijmolt, 2009). Estimating the models using alternative specifications of the matrix[12], I find that the highest log-likelihood function value is obtained using the first-order contiguity matrix for SDM and inverse distance matrix for DSDM (see Table 12 in Appendix C). Thus, I choose to estimate the models using the first-order contiguity matrix, a type of matrix that generally gives reliable estimates, better reveals the true spatial structure of the data and performs better on average than the others (Stakhovych & Bijmolt, 2009). However, all the log-likelihood values are very similar, so the estimation results should not be significantly impacted by the choice of spatial weight matrix, as suggested by LeSage and Pace (2009) (see Tables 13 and 14 in Appendix C for the estimations using alternative specifications of the spatial weight matrix).

3.3 Spatial heterogeneity

Educational attainment strongly vary between urban and rural areas (Rodríguez-Pose & Tselios, 2011). Moreover, Büchel and Van Ham (2003) showed that overeducation is affected by the size of the labour market, which is also related to urbanisation. Then, one can expect overeducation to be driven differently by the local context depending on the degree of urbanisation of the area. In particular, as economic activities are concentrated in urban areas, the difference in the structure of employment may be less important in these areas than in more rural ones. On the contrary, overeducation in the urban areas may be more affected by neighbouring labour markets as they attract more workers (Murphy et al., 1991).

[12] I use contiguity matrices (first- and second-order), k-nearest neighbours (from 3 to 7 neighbours) matrices and an inverse distance matrix (with a threshold at 100km, and a zero weight for areas located farther away).

Table 3: Specification tests for spatial and time dependence

Spatial weight matrix	Contiguity 1st-order	Contiguity 2nd-order	3NN	4NN	5NN	6NN	7NN	Inv. distance (max 100km)
(1) LM-Test for spatial lag								
Stat	69.00***	111.23***	52.00***	59.59***	78.34***	97.02***	105.45***	104.06***
p-value	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
(2) LM-Test for spatial error								
Stat	13.79***	14.95***	11.84***	13.49***	17.73***	21.51***	22.20***	25.34***
p-value	*0.000*	*0.000*	*0.001*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
(3) LR-Test for spatial lag of X (with lag of Y)								
Stat	60.67***	16.71	31.92***	40.02***	48.19***	36.61***	24.84**	46.38***
p-value	*0.000*	*0.161*	*0.001*	*0.000*	*0.000*	*0.000*	*0.016*	*0.000*
(4) LR-Test for time lag of Y								
Stat	62.68***	86.87***	54.03***	63.26***	64.68***	67.93***	69.25***	78.80***
p-value	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
(5) LR-Test for time lag and space-time lag of Y								
Stat	82.73***	87.03***	78.46***	86.00***	84.32***	80.27***	76.00***	87.75***
p-value	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
(6) LR-Test for space-time lag (with time lag of Y)								
Stat	20.04***	0.16	24.44***	22.73***	19.63***	12.34***	6.75***	8.96***
p-value	*0.000*	*0.688*	*0.000*	*0.000*	*0.000*	*0.000*	*0.009*	*0.003*

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

I choose to account for spatial heterogeneity by using spatial regime models (Le Gallo, 2004). The inclusion of spatial regimes allows the estimated effects of the explanatory variables to be different depending on the regime. The first regime corresponds to urban areas and the second regime to rural areas[13]. Following Le Gallo (2004), the model can be written as follows:

$$Y_t = \tau Y_{t-1} + \rho W Y_t + \eta W Y_{t-1} + D_U X_t \beta_U + W D_U X_t \theta_U + D_R X_t \beta_R + W D_R X_t \theta_R + \alpha + u_t$$

with D_U and D_R the dummy variables denoting the two spatial regimes, respectively equal to 1 if urban area (0 otherwise) and equal to 1 if rural area (0 otherwise). Such a model allows me to test for structural instability (equality of coefficients between regimes) through a Chow asymptotic test.

My model may suffer from an omitted variable bias, due to the absence of some information in the data. Indeed, there are some variables that are known to affect overeducation at the aggregate level that I cannot take into account: offered wages (Croce & Ghignoni, 2012), gross fixed capital formation (Ghignoni & Verashchagina, 2014) and GDP growth (Croce & Ghignoni, 2012), expenditures on students (Davia et al., 2017) or distance to big agglomerations (Büchel & Van Ham, 2003). However, the spatial specifications (Spatial Durbin models) that I use are more robust to omitted variables biases than non-spatial models (Pace & LeSage, 2010). Moreover, the spatial regimes models allow me to explore the impact of urbanisation, not included explicitly as an explanatory variable.

4 Results

4.1 Overall sample

Table 4 presents the estimation results for the fixed effects, SDM and DSDM models. Surprisingly, the coefficients for the spatially-lagged dependent variable in both SDM and DSDM are not significant. This means the situation in neighbouring areas impacts overeducation only through the explanatory variables. Indeed, contrary to what is observed for the dependent variable, the effects of several spatially-lagged regressors are significant. This confirms that the situation in neighbouring areas impacts overeducation. Since a spatial lag of the dependent variable is included in the SDM and DSDM, estimated coefficients do not directly give the im-

[13]Urban and rural areas are defined following the typology from Insee, which is based on population density. "Dense" and "intermediate density" areas are considered as urban, while "low density" and "very low density" areas are considered as rural. More details about this typology are provided in D'Alessandro et al. (2021).

pacts of the variables on overeducation. Direct, indirect and total effects for SDM and DSDM are reported in Table 5.

Table 4: Estimation results

Variables	**FE**	**SDM**		**DSDM**	
	X	X	WX	X	WX
% of 25-34 year-olds in the labour force	-0.041	-0.002	0.183***	-0.000	0.174***
	(0.039)	(0.027)	(0.059)	(0.025)	(0.055)
% of workers living outside their area of work	0.009	0.002	0.142***	-0.005	0.138***
	(0.033)	(0.023)	(0.054)	(0.022)	(0.050)
Number of microfirms per 100000 inhabitants (log)	0.032***	0.022***	0.011	0.021***	-0.001
	(0.004)	(0.006)	(0.008)	(0.005)	(0.008)
Number of newly established firms per 100000 inhabitants (log)	-0.014***	-0.010***	-0.016***	-0.010***	-0.010*
	(0.004)	(0.003)	(0.006)	(0.002)	(0.005)
% of employment in the agricultural sector	-0.060	-0.032	-0.104	-0.040	-0.014
	(0.070)	(0.048)	(0.110)	(0.045)	(0.104)
% of employment in the industrial sector	0.015	0.017	-0.104	0.008	-0.096
	(0.049)	(0.034)	(0.076)	(0.031)	(0.071)
% of employment in the construction sector	-0.099	-0.051	-0.198*	-0.063	-0.165
	(0.077)	(0.053)	(0.118)	(0.049)	(0.111)
% of employment in the sales sector	-0.015	-0.010	-0.206***	-0.007	-0.124*
	(0.048)	(0.034)	(0.073)	(0.032)	(0.070)
% of employment in the education, health and public sector	-0.109**	-0.115***	0.064	-0.125***	0.086
	(0.044)	(0.030)	(0.064)	(0.028)	(0.060)
Excess supply of tertiary graduates	0.123***	0.110***	-0.033	0.108***	-0.064***
	(0.011)	(0.008)	(0.021)	(0.008)	(0.021)
% of temporary contracts	0.025	0.029	-0.174***	0.031	0.175***
	(0.040)	(0.027)	(0.062)	(0.025)	(0.058)
Unemployment rate	0.039	-0.061	0.213**	-0.057	0.150*
	(0.060)	(0.043)	(0.086)	(0.040)	(0.081)
Lags of Y					
Spatial lag of Y		-0.051		0.047	
		(0.067)		(0.065)	
Time lag of Y				-0.039	
				(0.035)	
Space-time lag of Y				0.347***	
				(0.078)	
Poolability test Stat	5.70				
p-value	0.000				
Hausman test Stat	57.32				
p-value	0.000				
Log-likelihood	2701.98	2733.61		2774.97	
Observations	594	594		594	

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

I find similar effects in both the static and dynamic SDM specification. Moreover, the effects are significant both in the short and the long term. There is a strong direct effect of the imbalance between skilled labour supply and demand, confirming that the lack of qualified jobs in an area (relative to the number of tertiary graduates) leads to higher overeducation rates in this area. However, no direct effect of the unemployment rate is found on overeducation. This is in line with the results of Büchel and Van Ham (2003), who found that the labour markets where unemployment and overeducation occur do not share the same characteristics. These authors argued that different market mechanisms operate for overeducation and unemployment: the risk of unemployment is higher in more competitive labour markets, *i.e.* in areas where the unemployment rate is already high. By contrast, overeducation arises in narrower labour markets, *i.e.* in areas where there are less job opportunities, regardless of the characteristics

of the jobs available. Nevertheless, I find a positive indirect effect of unemployment rate. A one point increase in unemployment rate in the neighbouring areas leads to an increase of 0.2 points in the overeducation rate on average. This suggests that having unhealthy labour markets nearby promotes overeducation, although this seems to be a short-term mechanism given the results of the DSDM. In addition, the significant indirect negative effect of the number of newly established firms supports this view, with the proximity of a dynamic area reducing overeducation. These results also show a positive and significant indirect effect of the share of workers living outside their working area. This effect may indicate reverse causality: if there is overeducation in one area, workers are tempted to go to the neighbouring areas to find a more suitable job. In the DSDM specification, the effect of the space-time-lagged dependent variable is positive and highly significant. Where overeducation rates were one point higher in the neighbouring areas in the previous period (four years earlier), the overeducation rate is on average 0.35 points higher in the observed area in the current period. This may mean that some workers have moved from an area with a high overeducation rate in order to find a better job. According to findings by Langella and Manning (2022), if there is a lack of (suitable) job opportunities in an area, workers may be tempted to take a job in the neighbouring areas, overcrowding these labour markets and generating overeducation.

Surprisingly, I find a positive indirect effect of the share of 25-34 year-olds in the labour force. This means that the overeducation rate increases with the number of young workers in the neighbouring areas. Yet since young workers tend to live closer to job opportunities, in more dynamic areas, earlier results were suggestive of a negative sign. However, as young workers are more mobile, they may have relocated for better job opportunities. It would have been interesting to include information about individuals' previous residential migration to confirm this mechanism, but unfortunately, the data did not provide the necessary variables.

The number of microfirms has a positive effect on overeducation, which tends to confirm that smaller firms hire more overeducated workers, as highlighted by Ghignoni and Verashchagina (2014). Although I do not find significant effects of share of employment in most sectors, there is less overeducation in areas where the education, health and public sectors are more developed. This is in line with Meroni and Vera-Toscano (2017), who showed that the health sector is reasonably well preserved from overeducation. I also find a negative direct effect of the newly established firms in an area, confirming that economic dynamism reduces overeducation.

Table 5: Direct, indirect and total effects for SDM and DSDM models

		Short-term effects			**Long-term effects**		
	Variables	Direct	Indirect	Total	Direct	Indirect	Total
SDM	% of 25-34 year-olds in the labour force				-0.003	0.174***	0.171***
	% of workers living outside their area of work				-0.001	0.139***	0.138**
	Number of microfirms per 100000 inhabitants (log)				0.023***	0.009	0.031***
	Number of newly established firms per 100000 inhabitants (log)				-0.010***	-0.014***	-0.024***
	% of employment in the agricultural sector				-0.032	-0.083	-0.115
	% of employment in the industrial sector				0.020	-0.100	-0.080
	% of employment in the construction sector				-0.048	-0.187	-0.235*
	% of employment in the sales sector				-0.008	-0.197***	-0.205***
	% of employment in the education, health and public sector				-0.113***	0.066	-0.048
	Excess supply of tertiary graduates				0.110***	-0.037*	0.074***
	% of temporary contracts				0.030	-0.169***	-0.138**
	Unemployment rate				-0.061	0.209***	0.148*
DSDM	% of 25-34 year-olds in the labour force	0.004	0.177***	0.180***	0.016	0.251***	0.268***
	% of workers living outside their area of work	-0.005	0.142***	0.137**	0.006	0.198***	0.203**
	Number of microfirms per 100000 inhabitants (log)	0.021***	0.000	0.021***	0.020***	0.010	0.031***
	Number of newly established firms per 100000 inhabitants (log)	-0.010***	-0.011*	-0.021***	-0.011***	-0.020***	-0.031***
	% of employment in the agricultural sector	-0.040	-0.017	-0.056	-0.041	-0.043	-0.083
	% of employment in the industrial sector	0.007	-0.096	-0.089	-0.000	-0.132	-0.132
	% of employment in the construction sector	-0.062	-0.176	-0.238**	-0.074	-0.279*	-0.353*
	% of employment in the sales sector	-0.008	-0.128*	-0.136*	-0.017	-0.184*	-0.201*
	% of employment in the education, health and public sector	-0.125***	0.082	-0.043	-0.118***	0.055	-0.063
	Excess supply of tertiary graduates	0.108***	-0.063***	0.045**	0.102***	-0.036	0.066**
	% of temporary contracts	0.028	-0.182***	-0.154**	0.014	-0.243***	-0.229***
	Unemployment rate	-0.054	0.154*	0.100	-0.042	0.190	0.148

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

4.2 Subsamples by gender

Tables 6 and 7 respectively present the results for the female and male subsamples. I estimate both the static and dynamic SDM models using the same spatial weight matrix as for the main analysis, *i.e.* the first-order contiguity matrix. This ensures the comparability with the results of the main analysis, and is also consistent with the log-likelihoods obtained with different specifications of the matrix (see Table 12 in Appendix C).

Overall, results for women are similar to those obtained with the main sample. Nevertheless, there are a few effects that are specific to the female subsample. First, the distribution of employment by sector, both in the area and in the neighbouring areas, appears to have a larger impact on the overeducation of women than on overall overeducation. Indeed, almost every sector is associated with a significant coefficient in this specification, whereas the education, health and the public sector showed a significant effect in the previous model. Then, a significant positive effect of the share of workers living outside the area is observed on women's overeducation, while this effect is negative for the male sample and not significant at all for the main specification. This suggests that competition for skilled jobs has a greater impact on women than on the other demographic groups: women are more likely to be pushed down into less qualified jobs when there is a shortage of more qualified positions.

The effects on men's overeducation are different from those observed in the overall sample. There is no longer a positive indirect effect from the unemployment rate, but rather a negative direct effect. This means that men's overeducation rates are lower in areas with higher unemployment. This may reflect, at the local level, the trade-off between unemployment and overeducation, as described by Simpson (1992): individuals can choose to remain unemployed to avoid overeducation. Finally, one effect is opposite to that observed for women: the overeducation rate for men is lower in more competitive labour markets, with negative direct effects found for the share of 25-34 year-olds (on average more educated) and the share of workers living outside the employment area.

Table 6: Estimation results and direct, indirect and total effects for SDM and DSDM - Female subsample

		Coefficients		**Short-term effects**			**Long-term effects**		
	Variables	X	WX	Direct	Indirect	Total	Direct	Indirect	Total
SDM	% of 25-34 year-olds in the labour force	0.033	0.236***				0.033	0.224***	0.257***
	% of workers living outside their area of work	0.114***	0.279***				0.111***	0.271***	0.382***
	Number of microfirms per 100000 inhabitants (log)	0.036***	0.006				0.037***	0.003	0.040**
	Number of newly established firms per 100000 inhabitants (log)	-0.009**	-0.017**				-0.010**	-0.016*	-0.026***
	% of employment in the agricultural sector	-0.103	-0.148				-0.103	-0.117	-0.220
	% of employment in the industrial sector	0.102**	-0.105				0.105**	-0.105	0.000
	% of employment in the construction sector	0.069	-0.357**				0.073	-0.347*	-0.273
	% of employment in the sales sector	0.034	-0.453***				0.037	-0.440***	-0.403***
	% of employment in the education, health and public sector	-0.147***	0.125				-0.144***	0.124	-0.020
	Excess supply of tertiary graduates	0.114***	-0.094***				0.116***	-0.095***	0.020
	% of temporary contracts	0.052	-0.307***				0.054	-0.298***	-0.244**
	Unemployment rate	0.008	0.179				0.009	0.176	0.186
	Spatial lag of Y	-0.041							
	Log-likelihood: 2489.73								
DSDM	% of 25-34 year-olds in the labour force	0.040	0.239***	0.045	0.240***	0.286***	0.057	0.321***	0.378***
	% of workers living outside their area of work	0.100***	0.251***	0.101***	0.260***	0.360***	0.113***	0.364***	0.477***
	Number of microfirms per 100000 inhabitants (log)	0.034***	-0.001	0.034***	0.001	0.035***	0.033**	0.013	0.046***
	Number of newly established firms per 100000 inhabitants (log)	-0.010***	-0.013*	-0.010***	-0.014*	-0.024***	-0.010***	-0.022**	-0.032***
	% of employment in the agricultural sector	-0.098	-0.082	-0.099	-0.088	-0.187	-0.101	-0.145	-0.247
	% of employment in the industrial sector	0.092*	-0.099	0.090*	-0.093	-0.003	0.083*	-0.087	-0.004
	% of employment in the construction sector	0.061	-0.331**	0.063	-0.339**	-0.276	0.043	-0.409*	-0.366
	% of employment in the sales sector	0.034	-0.394***	0.033	-0.401***	-0.368***	0.011	-0.498***	-0.487***
	% of employment in the education, health and public sector	-0.154***	0.149*	-0.154***	0.146*	-0.008	-0.143***	0.133	-0.010
	Excess supply of tertiary graduates	0.112***	-0.110***	0.111***	-0.111***	0.000	0.103***	- 0.103***	0.000
	% of temporary contracts	0.052	-0.318***	0.048	-0.327***	-0.279***	0.029	-0.399***	0.370***
	Unemployment rate	0.014	0.132	0.017	0.136	0.153	0.024	0.178	0.203
	Spatial lag of Y	0.030							
	Time lag of Y	-0.040							
	Space-time lag of Y	0.275***							
	Log-likelihood: 2533.36								

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

Table 7: Estimation results and direct, indirect and total effects for SDM and DSDM - Male subsample

		Coefficients		**Short-term effects**			**Long-term effects**		
	Variables	X	WX	Direct	Indirect	Total	Direct	Indirect	Total
SDM	% of 25-34 year-olds in the labour force	-0.107***	0.139*				-0.105***	0.135*	0.030
	% of workers living outside their area of work	-0.093***	-0.039				-0.094***	-0.039	-0.134*
	Number of microfirms per 100000 inhabitants (log)	0.008	0.007				0.009	0.007	0.015**
	Number of newly established firms per 100000 inhabitants (log)	-0.011***	-0.011				-0.011***	-0.011	-0.022***
	% of employment in the agricultural sector	-0.027	-0.208				-0.028	-0.192	-0.221
	% of employment in the industrial sector	-0.046	-0.209**				-0.044	-0.214**	-0.258**
	% of employment in the construction sector	-0.163**	-0.099				-0.162**	-0.103	-0.265
	% of employment in the sales sector	-0.063	0.012				-0.063	0.013	-0.050
	% of employment in the education, health and public sector	-0.122***	-0.022				-0.119***	-0.028	-0.147*
	Excess supply of tertiary graduates	0.098***	0.029				0.098***	0.032	0.131***
	% of temporary contracts	-0.004	-0.058				-0.005	-0.059	-0.064
	Unemployment rate	-0.166***	0.182				-0.163***	0.184*	0.021
	Spatial lag of Y	0.020							
	Log-likelihood: 2577.47								
DSDM	% of 25-34 year-olds in the labour force	-0.114***	0.186**	-0.110***	0.181**	0.071	-0.141***	0.245**	0.103
	% of workers living outside their area of work	-0.099***	-0.042	-0.099***	-0.045	-0.144*	-0.133***	-0.078	-0.210*
	Number of microfirms per 100000 inhabitants (log)	0.005	0.003	0.005	0.004	0.008	0.006	0.006	0.012
	Number of newly established firms per 100000 inhabitants (log)	-0.011***	-0.011	-0.011***	-0.012	-0.022***	-0.014***	-0.018*	-0.032***
	% of employment in the agricultural sector	-0.005	-0.266*	-0.006	-0.270*	-0.276*	-0.013	-0.388*	-0.401*
	% of employment in the industrial sector	-0.064	-0.271***	-0.067	-0.272***	-0.338***	-0.094	-0.398***	-0.492***
	% of employment in the construction sector	-0.195***	-0.158	-0.192***	-0.165	-0.357**	-0.258***	-0.262	-0.520**
	% of employment in the sales sector	-0.050	0.032	-0.048	0.034	-0.014	-0.063	0.043	-0.020
	% of employment in the education, health and public sector	-0.120***	-0.032	-0.121***	-0.036	-0.014	-0.160***	-0.068	-0.228**
	Excess supply of tertiary graduates	0.086***	0.020	0.086***	0.020	0.107***	0.114***	0.040	0.155***
	% of temporary contracts	-0.015	-0.074	-0.017	-0.077	-0.095	-0.025	-0.114	-0.139
	Unemployment rate	-0.202***	0.104	-0.199***	0.101	-0.098	-0.261***	0.118	-0.143
	Spatial lag of Y	0.016							
	Time lag of Y	0.242***							
	Space-time lag of Y	0.063							
	Log-likelihood: 2591.04								

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

4.3 Youth subsample

Table 8 displays the results for the 25-34 year-olds. Most indirect effects are not significant, suggesting that this demographic group is less impacted by the situation in neighbouring areas than the whole population. This is not surprising: younger workers are more mobile, having fewer household responsibilities. If there is a shortage of suitable jobs available on a local labour market, they are thus likely to move to other markets with more opportunities, located farther away from the less dynamic areas.

I find that overeducation rates decrease as the share of young workers in the labour force increases. At first sight, this may seem counterintuitive: because of their higher level of education, more young workers should lead to more competition for skilled jobs, and thus more overeducation. However, as mentioned earlier, young workers are not randomly located. On average, they live in bigger agglomerations, with more economic dynamism and more job opportunities, which may make it easier for them to find a suitable job in these employment areas.

The share of temporary contracts has a negative direct effect, whereas there was no significant effect in the overall sample. This suggests that, for young workers especially, repeated job seeking can be a way to improve matches (Jovanovic, 1979), and that using temporary jobs (well-matched, but with less job security) as a "stepping-stone" at the beginning of a career (Ghignoni & Verashchagina, 2014) can lead to reduced educational mismatch.

Table 8: Estimation results and direct, indirect and total effects for SDM and DSDM - Youth subsample

		Coefficients		**Short-term effects**			**Long-term effects**		
	Variables	X	WX	Direct	Indirect	Total	Direct	Indirect	Total
SDM	% of 25-34 year-olds in the labour force	-0.127*	0.221				-0.130*	0.207	0.078
	% of workers living outside their area of work	0.071	0.238*				0.061	0.211	0.273*
	Number of microfirms per 100000 inhabitants (log)	0.019	-0.009				0.021	-0.013	0.007
	Number of newly established firms per 100000 inhabitants (log)	-0.013*	0.025*				-0.014**	0.025*	0.011
	% of employment in the agricultural sector	-0.186	0.157				-0.192	0.203	0.010
	% of employment in the industrial sector	0.110	-0.409**				0.127	-0.383**	-0.256
	% of employment in the construction sector	-0.014	-0.352				-0.001	-0.311	-0.313
	% of employment in the sales sector	0.071	-0.260				0.078	-0.241	-0.163
	% of employment in the education, health and public sector	-0.087	0.123				-0.084	0.116	0.032
	Excess supply of tertiary graduates	0.141***	-0.019				0.143***	-0.034	0.109**
	% of temporary contracts	-0.147**	-0.338**				-0.140*	-0.283*	-0.423***
	Unemployment rate	-0.087	-0.114				-0.079	-0.084	-0.163
	Spatial lag of Y	-0.145**							
	Log-likelihood: 2149.23								
DSDM	% of 25-34 year-olds in the labour force	-0.113	0.220	-0.111*	0.200	0.089	-0.141	0.258	0.118
	% of workers living outside their area of work	0.037	0.228	0.030	0.200	0.230*	0.039	0.263	0.302*
	Number of microfirms per 100000 inhabitants (log)	0.034**	-0.022	0.035**	-0.024	0.012	0.045**	-0.030	0.015
	Number of newly established firms per 100000 inhabitants (log)	-0.012*	0.027*	-0.012*	0.025*	0.013	-0.016*	0.033*	0.017
	% of employment in the agricultural sector	-0.187	0.235	-0.194	0.237	0.043	-0.246	0.304	0.058
	% of employment in the industrial sector	0.138	-0.383*	0.145	-0.353**	-0.209	0.183	-0.457*	-0.274
	% of employment in the construction sector	-0.018	-0.432	0.000	-0.392	-0.391	-0.002	-0.513	-0.515
	% of employment in the sales sector	0.082	-0.185	0.089	-0.173	-0.085	0.112	-0.223	-0.111
	% of employment in the education, health and public sector	-0.086	0.134	-0.091	0.129	0.038	-0.116	0.166	0.050
	Excess supply of tertiary graduates	0.131***	-0.051	0.134***	-0.066	0.067	0.171***	-0.082	0.089
	% of temporary contracts	-0.144**	-0.284*	-0.141**	-0.239	-0.381**	-0.183**	-0.318	-0.501**
	Unemployment rate	-0.131	-0.049	-0.125	-0.027	-0.152	-0.161	-0.039	-0.200
	Spatial lag of Y	0.139*							
	Time lag of Y	0.220***							
	Space-time lag of Y	0.052							
	Log-likelihood: 2164.84								

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

4.4 Heterogeneity between urban and rural areas

Estimation results for the spatial regimes models are reported in Table 9. The results of the Chow asymptotic test for structural instability confirm that overducation is not driven by the same factors in urban and rural areas. Coefficients associated with the lagged dependent variable (both in space and time dimensions) are similar to the ones obtained through the models without heterogeneity. However, the spatial lag of the dependent variable is now significant (although at the 10% level) in the spatial regimes SDM, while it was not in the previous specification. This means overeducation rates tend to be lower when more individuals are mismatched in neighbouring areas.

Tables 10 and 11 respectively display the direct, indirect and total effects for the SDM and DSDM with spatial regimes. The results observed for rural areas are similar to the ones obtained without taking spatial heterogeneity into account, in both the static and the dynamic specifications. It has to be noted that the structure of employment by sector have significant effects in rural areas, while it does not seem to be the case in urban areas. Spatial regimes models confirm the positive direct effect of the excess supply of tertiary graduates and the number of microfirms (both for urban and rural areas) and the negative direct effect of economic dynamism (proxied by the number of newly established firms) on overeducation rates.

However, most results for urban areas differ from the models without heterogeneity. The share of workers living outside their working area now has a negative (although barely significant at the 10% level) effect on overeducation. This may reveal that workers come to cities to find more suitable jobs than in their area of residence. I still find a positive negative indirect effect of the unemployment rate (only in the short term for the dynamic specification), however, I also find a negative direct effect of that same variable in both the SDM and DSDM specifications. Urban areas are more attractive for workers, resulting in more competition for jobs, but also more opportunities. Therefore, workers may accept a job for which they are overeducated in order to quit unemployment more quickly, with the expectation to find a more adequate position through on-the-job search or promotion. Young workers are more prone to adopt this behaviour (Sicherman & Galor, 1990), which can explain the positive effect of the share of 25-34 year-olds, even though Baert et al. (2013) showed that this strategy can lead to "overeducation traps".

Table 9: Estimation results for urban and rural areas - Spatial regimes models

	SDM				DSDM			
	Urban areas		Rural areas		Urban areas		Rural areas	
Variables	X	WX	X	WX	X	WX	X	WX
% of 25-34 year-olds in the labour force	0.135**	0.087	-0.027	0.282***	0.145**	0.076	-0.017	0.265***
	(0.062)	(0.142)	(0.029)	(0.066)	(0.058)	(0.132)	(0.027)	(0.062)
% of workers living outside their area of work	0.004	-0.214*	0.006	0.219***	0.002	-0.218*	-0.010	0.202***
	(0.049)	(0.123)	(0.025)	(0.063)	(0.045)	(0.113)	(0.023)	(0.058)
Number of microfirms per 100000 inhabitants (log)	0.030***	0.017	0.026***	0.010	0.026***	0.007	0.023***	-0.007
	(0.008)	(0.014)	(0.006)	(0.010)	(0.007)	(0.013)	(0.006)	(0.010)
Number of newly established firms per 100000 inhabitants (log)	-0.023***	-0.056***	-0.007***	-0.005	-0.023***	-0.053***	-0.008***	0.002
	(0.006)	(0.013)	(0.003)	(0.007)	(0.005)	(0.012)	(0.002)	(0.006)
% of employment in the agricultural sector	-0.108	0.371	-0.071	-0.175	-0.142	0.507*	-0.088*	-0.045
	(0.149)	(0.320)	(0.051)	(0.121)	(0.137)	(0.296)	(0.047)	(0.113)
% of employment in the industrial sector	0.015	-0.153	-0.013	-0.107	-0.033	-0.226*	-0.020	-0.075
	(0.069)	(0.144)	(0.038)	(0.090)	(0.064)	(0.134)	(0.035)	(0.083)
% of employment in the construction sector	0.207*	0.117	-0.071	-0.337**	0.250**	0.101	-0.106*	-0.298**
	(0.111)	(0.235)	(0.060)	(0.145)	(0.103)	(0.220)	(0.055)	(0.135)
% of employment in the sales sector	0.206***	-0.183	-0.098**	-0.203**	0.227***	0.012	-0.097***	-0.123
	(0.063)	(0.160)	(0.040)	(0.086)	(0.058)	(0.152)	(0.037)	(0.080)
% of employment in the education, health and public sector	-0.013	0.037	-0.155***	-0.013	-0.005	0.076	-0.187***	-0.017
	(0.054)	(0.112)	(0.035)	(0.083)	(0.050)	(0.104)	(0.033)	(0.077)
Excess supply of tertiary graduates	0.125***	-0.093**	0.112***	-0.017	0.119***	-0.191***	0.111***	-0.050**
	(0.019)	(0.043)	(0.008)	(0.022)	(0.019)	(0.042)	(0.008)	(0.021)
% of temporary contracts	-0.006	0.065	0.018	-0.199***	-0.004	0.031	0.017	-0.208***
	(0.061)	(0.145)	(0.030)	(0.069)	(0.056)	(0.134)	(0.028)	(0.064)
Unemployment rate	-0.192***	0.318*	-0.018	0.287**	-0.175***	0.281*	-0.006	0.152
	(0.072)	(0.163)	(0.051)	(0.117)	(0.066)	(0.151)	(0.047)	(0.111)
Spatial lag of Y			-0.132*				0.012	
			(0.069)				(0.067)	
Time lag of Y							-0.044	
							(0.035)	
Space-time lag of Y							0.512***	
							(0.081)	
Chow test for structural instability			80.43				107.38	
			(0.000)				(0.000)	
Log-likelihood			2769.33				2817.72	
Observations	206		388		206		388	

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

Table 10: Direct, indirect and total effects for spatial regime SDM

		Long-term effects		
	Variables	Direct	Indirect	Total
Urban areas	% of 25-34 year-olds in the labour force	0.135**	0.059	0.194
	% of workers living outside their area of work	0.014	-0.194*	-0.180
	Number of microfirms per 100000 inhabitants (log)	0.029***	0.012	0.041***
	Number of newly established firms per 100000 inhabitants (log)	-0.021***	-0.048***	-0.069***
	% of employment in the agricultural sector	-0.104	0.354	0.249
	% of employment in the industrial sector	0.016	-0.138	-0.122
	% of employment in the construction sector	0.201*	0.068	0.269
	% of employment in the sales sector	0.209***	-0.189	0.020
	% of employment in the education, health and public sector	-0.010	0.027	0.018
	Excess supply of tertiary graduates	0.127***	-0.099**	0.028
	% of temporary contracts	-0.013	0.060	0.047
	Unemployment rate	-0.202***	0.303**	0.101
Rural areas	% of 25-34 year-olds in the labour force	-0.035	0.259***	0.224***
	% of workers living outside their area of work	0.000	0.200***	0.201***
	Number of microfirms per 100000 inhabitants (log)	0.027***	0.006	0.033***
	Number of newly established firms per 100000 inhabitants (log)	-0.007***	-0.004	-0.011*
	% of employment in the agricultural sector	-0.064	-0.150	-0.214*
	% of employment in the industrial sector	-0.007	-0.091	-0.098
	% of employment in the construction sector	-0.058	-0.295**	-0.352***
	% of employment in the sales sector	-0.091**	-0.168**	-0.259***
	% of employment in the education, health and public sector	-0.155***	0.006	-0.148*
	Excess supply of tertiary graduates	0.113***	-0.031*	0.082***
	% of temporary contracts	0.024	-0.184***	-0.160**
	Unemployment rate	-0.026	0.260**	0.234**

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

Table 11: Direct, indirect and total effects for spatial regime DSDM

		Short-term effects			**Long-term effects**		
	Variables	Direct	Indirect	Total	Direct	Indirect	Total
Urban areas	% of 25-34 year-olds in the labour force	0.151***	0.076	0.227*	0.162***	0.262	0.424*
	% of workers living outside their area of work	0.003	-0.213*	-0.210*	-0.022	-0.373*	-0.395*
	Number of microfirms per 100000 inhabitants (log)	0.027***	0.006	0.033***	0.028***	0.034*	0.062***
	Number of newly established firms per 100000 inhabitants (log)	-0.023***	-0.052***	-0.075***	-0.029***	-0.112***	-0.141***
	% of employment in the agricultural sector	-0.146	0.491*	0.345	-0.090	0.744	0.653
	% of employment in the industrial sector	-0.038	-0.239*	-0.276*	-0.066	-0.454*	-0.520*
	% of employment in the construction sector	0.241**	0.093	0.333	0.255**	0.373	0.628
	% of employment in the sales sector	0.229***	-0.002	0.226	0.232***	0.195	0.427
	% of employment in the education, health and public sector	-0.005	0.075	0.070	0.003	0.129	0.133
	Excess supply of tertiary graduates	0.119***	-0.190***	-0.072*	0.098***	-0.234***	-0.136
	% of temporary contracts	-0.005	0.020	0.015	-0.003	0.031	0.028
	Unemployment rate	-0.180***	0.282*	0.102	-0.149**	0.342	0.193
Rural areas	% of 25-34 year-olds in the labour force	-0.018	0.263***	0.245***	0.012	0.448***	0.460***
	% of workers living outside their area of work	-0.009	0.211***	0.202***	0.016	0.364***	0.380***
	Number of microfirms per 100000 inhabitants (log)	0.023***	-0.007	0.016*	0.023***	0.007	0.030*
	Number of newly established firms per 100000 inhabitants (log)	-0.008***	0.002	-0.006	-0.008***	-0.003	-0.010
	% of employment in the agricultural sector	-0.084*	-0.042	-0.126	-0.090*	-0.147	-0.237
	% of employment in the industrial sector	-0.017	-0.072	-0.090	-0.026	-0.143	-0.167
	% of employment in the construction sector	-0.103*	-0.292**	-0.394***	-0.139**	-0.603**	-0.741**
	% of employment in the sales sector	-0.096**	-0.120	-0.216***	-0.111***	-0.293**	-0.404***
	% of employment in the education, health and public sector	-0.185***	-0.012	-0.197**	-0.189***	-0.181	-0.370**
	Excess supply of tertiary graduates	0.111***	-0.051**	0.061***	0.107***	0.006	0.113***
	% of temporary contracts	0.017	-0.207***	-0.190***	-0.007	-0.350***	-0.357***
	Unemployment rate	-0.008	0.155	0.148	0.011	0.267	0.278

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

5 Conclusion

The aim of this work was to study the local determinants of overeducation and the disparities between local overeducation rates in metropolitan France in the 2010s. To this end, I regress overeducation rates on a large set of local characteristics for each French employment area. To take the spatial structure of the data into account and to explore how neighbouring labour markets can impact each other, I estimate static and dynamic spatial econometrics models that account for time dependency along with spatial dependency. I also estimate spatial regimes models to take into account the spatial heterogeneity between urban and rural areas.

Unlike previous studies (Croce & Ghignoni, 2012; Ghignoni & Verashchagina, 2014), I find a significant effect of the excess supply of skilled labour in the long term. However, I show that this is neither the sole nor the main driver of educational mismatch. First, the geography of labour markets is identified as an important driver of overeducation, with a strong spatial dependence found between labour markets: the situation in neighbouring areas has significant effects. In particular, I find that overeducation is higher in areas surrounded by less dynamic labour markets. My results also suggest that the inter-area mobility of workers impacts overeducation, particularly for women, where the share of workers coming from another area has a strong effect. This also confirms that the competition for skilled jobs has a greater effect on women, who are more at risk of being pushed down into less skilled jobs than other workers. By contrast, men are less overeducated on more competitive labour markets. As for young workers, they seem to be less affected by the geography of local labour markets, which can be explained by their higher spatial mobility. The results for spatial regimes models confirm that overeducation local drivers vary between urban and rural areas. While overeducation is affected by the structure of employment by sector in rural areas, it appears not to be the case in urban areas. On the contrary, I find strong effects of the share of young workers and of the unemployment rate in urban areas, that are not significant in rural areas. This suggests that overeducation is driven by (the lack of) opportunities in rural areas, and by competition in urban areas.

The results show that scarcity of job opportunities in one area is likely to increase educational mismatch in neighbouring areas. From a public policy point of view, this suggests that local employment policies may well improve the labour market at a broader scale than in their particular area of implementation. In particular, increasing the dynamism of rural areas, where workers experience a lack of job opportunities, might reduce overeducation in these areas, but also in urban areas, as mismatched workers from rural areas would be less likely to go to cities to find a suitable job, and thus to overcrowd already competitive labour markets.

References

Baert, S., Cockx, B., & Verhaest, D. (2013). Overeducation at the start of the career: Stepping stone or trap? *Labour Economics*, *25*, 123–140.

Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New York: Columbia University Press.

Bouayad Agha, S., Le Gallo, J., & Vedrine, L. (2018). Econométrie spatiale sur données de panel. In *Manuel d'analyse spatiale, Insee Méthodes 131*. INSEE.

Büchel, F., & Van Ham, M. (2003). Overeducation, regional labor markets,and spatial flexibility. *Journal of Urban Economics*, *53*(3), 482–493.

Cliff, A. D., & Ord, J. K. (1973). *Spatial Autocorrelation*. London: Pion.

Cliff, A. D., & Ord, J. K. (1981). *Spatial processes, Models and Applications*. London: Pion.

Croce, G., & Ghignoni, E. (2012). Demand and supply of skilled labour and overeducation in Europe: A country-level analysis. *Comparative Economic Studies*, *54*(2), 413–439.

D'Alessandro, C., Levy, D., & Regnier, T. (2021). Une nouvelle définition du rural pour mieux rendre compte des réalités des territoires et de leurs transformations. In *La France et ses territoires, Insee References*. INSEE.

Davia, M. A., McGuinness, S., & O'Connell, P. J. (2017). Determinants of regional differences in rates of overeducation in Europe. *Social Science Research*, *63*, 67–80.

Debarsy, N., & Ertur, C. (2010). Testing for spatial autocorrelation in a fixed effects panel data model. *Regional Science and Urban Economics*, *40*(6), 453–470.

Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International regional science review*, *26*(3), 244–268.

Elhorst, J. P. (2014). Spatial panel data models. In *Spatial econometrics* (pp. 37–93). Springer, Berlin, Heidelberg.

Frank, R. (1978). Why women earn less: The theory and estimation of differential overqualification. *American Economic Review*, *68*(3), 360–373.

Ghignoni, E., & Verashchagina, A. (2014). Educational qualifications mismatch in Europe. Is it demand or supply driven? *Journal of Comparative Economics*, *42*(3), 670–692.

Groot, W., & Maassen van den Brink, H. (2000). Overeducation in the labour market: A meta-analysis. *Economics of Education Review*, *21*(2), 149–158.

Guironnet, J.-P. (2006). La suréducation en france : vers une dévalorisation des diplômes du supérieur ? *Economie appliquée*, *59*(1), 93–120.

Hensen, M. M., De Vries, M. R., & Cörvers, F. (2009). The role of geographic mobility in reducing education-job mismatches in the Netherlands. *Papers in Regional Science*, *88*(3), 667–682.

Jauhiainen, S. (2011). Overeducation in the Finnish regional labour markets. *Papers in Regional Science*, *90*(3), 573–588.

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy*, *87*(5, Part 1), 972–990.

Kucel, A. (2011). Literature survey of the incidence of over-education: A sociological approach. *Revista Española de Investigationes Sociologicas*, *1*(134), 125–142.

Langella, M., & Manning, A. (2022). Residential mobility and unemployment in the UK. *Labour Economics*, *75*, 102104.

Le Gallo, J. (2004). Hétérogénéité spatiale, principes et méthodes. *Economie & prévision*, *162*(1), 151–172.

LeSage, J. P., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. CRC Press.

Manning, A., & Petrongolo, B. (2017). How local are labor markets? Evidence from a spatial job search model. *American Economic Review*, *107*(10), 2877–2907.

Marinescu, I., & Rathelot, R. (2018). Mismatch unemployment and the geography of job search. *American Economic Journal: Macroeconomics*, *10*(3), 42–70.

Martin, P., & Ottaviano, G. I. (2001). Growth and agglomeration. *International Economic Review*, *42*(4), 947–968.

McGuinness, S. (2006). Overeducation in the labour market. *Journal of Economic Surveys*, *20*(3), 387–418.

Meroni, E. C., & Vera-Toscano, E. (2017). The persistence of overeducation among recent graduates. *Labour Economics*, *48*, 120–143.

Moran, P. A. (1950). A test for serial independence of residuals. *Biometrika*, *37*(1/2), 178–181.

Murphy, K. M., Shleifer, A., & Vishny, R. W. (1991). The allocation of talent: Implications for growth. *The Quarterly Journal of Economics*, *106*(2), 503–530.

Nauze-Fichet, E., & Tomasini, M. (2002). Diplôme et insertion sur le marché du travail : Approches socioprofessionnelle et salariale du déclassement suivi d'un commentaire de Saïd Hanchane et Eric Verdier. *Economie et statistique*, *354*(1), 21–44.

Pace, R. K., & LeSage, J. P. (2010). Omitted variable biases of ols and spatial lag models. In *Progress in spatial analysis* (pp. 17–28). Springer, Berlin, Heidelberg.

Patacchini, E., & Zenou, Y. (2007). Spatial dependence in local unemployment rates. *Journal of Economic Geography*, *7*(2), 169–191.

Rodríguez-Pose, A., & Tselios, V. (2011). Mapping the European regional educational distribution. *European Urban and Regional Studies*, *18*(4), 358–374.

Rodríguez-Pose, A., & Vilalta-Bufí, M. (2005). Education, migration, and job satisfaction: The regional returns of human capital in the EU. *Journal of Economic Geography*, *5*(5), 545–566.

Sattinger, M. (1993). Assignment models of the distribution of earnings. *Journal of Economic Literature*, *31*, 831–880.

Sicherman, N., & Galor, O. (1990). A theory of career mobility. *Journal of Political Economy*, *98*(1), 169–192.

Simpson, W. (1992). *Urban structure and the labour market. Worker mobility, commuting and underemployment in cities*. Clarendon Press, Oxford.

Sloane, P., Battu, H., & Seaman, P. (1999). Overeducation, undereducation and the British labour market. *Applied Economics*, *31*, 1437–1453.

Stakhovych, S., & Bijmolt, T. H. (2009). Specification of spatial models: A simulation study on weights matrices. *Papers in Regional Science*, *88*(2), 389–408.

Thurow, L. C. (1975). *Generating inequality*. Basic Books: New York.

Tsang, M., Rumberger, R., & Levin, H. (1991). The impact of surplus scooling on worker productivity. *Industrial Relations*, *30*(2), 209–228.

Tselios, V. (2013). Overeducation and externalities in the EU: The combined moderating influence of migration and gender. *Environment and Planning C: Government and Policy*, *31*(2), 193–221.

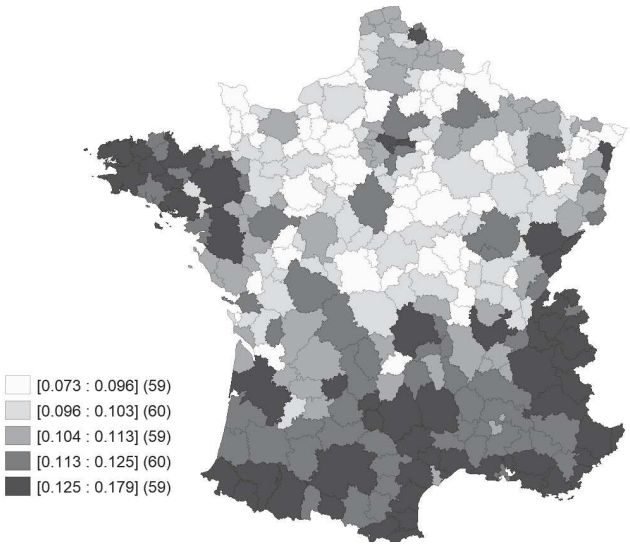
Verdugo, R. R., & Verdugo, N. T. (1989). The impact of surplus schooling on earnings: Some additional findings. *Journal of Human Resources*, 629–643.

Appendix A: Geography of overeducation using an alternative measure

To check the robustness of my measure, I calculate overeducation rates using an alternative measure, based on the median and the quartiles of the distribution. The idea behind this measure is that there is an educational mismatch for the degrees farthest from the median. The structure of the data allows me to consider seven levels of education : (1) No degree, (2) 2-year vocational high school, (3) 3-year vocational high school, (4) general high school, (5) 2-year tertiary degree, (6) 3- and 4-year tertiary degree and (7) 5-year tertiary degree and higher. For each of the professional families, individuals are overeducated if their level of schooling is higher than the third quartile of the distribution of education for this particular family.

Overeducation rates by employment area are represented in Figure 3. As expected, overeducation rates are higher using this alternative measure of overeducation. Workers possessing a more advanced degree but working in an occupation requiring more basic tertiary education are now considered overeducated, whereas they were not identified as overeducated by the measure used in the main analysis. Nevertheless, the spatial distribution of overeducation does not substantially vary between the two measures (compare Figures 2 and 3). Moreover, there is a high correlation coefficient between the two measures of overeducation (around 0.87). Therefore, my results should not be significantly impacted by the lack of precision of the main measure.

Figure 3: Overeducation rates by French employment area - 2017 - Alternative measure



Note: Areas in white have the lowest overeducation rates, areas in dark grey have the highest overeducation rates.

Appendix B: Direct, indirect and total effects in DSDM and SDM models

The Dynamic Spatial Durbin Model (DSDM) includes a time-lagged dependent variable Y_{t-1}, a spatially lagged dependent variable WY_t, a space-time-lagged dependent WY_{t-1} and spatially lagged independent variables WX_t. Then, the model can be written as follows:

$$Y_t = \tau Y_{t-1} + \rho W Y_t + \eta W Y_{t-1} + X_t\beta + WX_t\theta + \alpha + u_t$$

As the model includes WY, the coefficients β and θ associated with independent variables cannot be interpreted directly. Indeed, while a change in X in an area directly impacts Y in this particular area, it also impacts Y in the other areas through W. To identify direct and indirect effects, I rewrite the model as follows:

$$Y_t = (I-\rho W)^{-1}(\tau I + \eta W)Y_{t-1} + (I-\rho W)^{-1}X_t\beta + (I-\rho W)^{-1}WX_t\theta + (I-\rho W)^{-1}(\alpha + u_t)$$

$$Y_t = (I-\rho W)^{-1}(\tau I + \eta W)Y_{t-1} + (I-\rho W)^{-1}(\beta I_N + \theta W)X_t + (I-\rho W)^{-1}(\alpha + u_t)$$

$$Y_t = \sum_{k=1}^{N}(I-\rho W)^{-1}(\tau I + \eta W)Y_{t-1} + (I-\rho W)^{-1}(\beta_k I_N + \theta_k W)X_{kt} + (I-\rho W)^{-1}(\alpha + u_t)$$

where k denotes the kth explanatory variable.

The short term effects of the kth explanatory variable are then given by the matrix of partial derivatives of $E(Y)$ with respect to the kth variable in areas 1 to N at a particular point of time:

$$\left[\frac{\partial E(Y)}{\partial x_{1k}} \cdots \frac{\partial E(Y)}{\partial x_{Nk}}\right]_t = (I-\rho W)^{-1}(\beta_k I_N + \theta_k W)$$

Each diagonal element of this matrix is the short-term direct effect for an area of a change in variable X_k in this particular area. Thus, the average short-term direct effect corresponds to the mean diagonal element of this matrix, *i.e.* the average of the diagonal terms. The other elements of this matrix are the short-term indirect effects of a change in X_k in a given area on Y in the other areas. Thus, the average short-term indirect effect is given by the mean row sum of the non-diagonal elements. The average short-term total effect is the sum of the average short-term direct effect and the average short-term indirect effect.

Similarly, the long-term effects of the kth explanatory variable are given by:

$$\left[\frac{\partial E(Y)}{\partial x_{1k}} \cdots \frac{\partial E(Y)}{\partial x_{Nk}}\right] = [(1-\tau)I - (\rho+\eta)W]^{-1}(\beta_k I_N + \theta_k W)$$

where the average long-term direct effect is given by the mean diagonal element of the matrix and the average long-term indirect effect is given by the mean row sum of the non-diagonal elements.

The SDM model also distinguishes between direct and indirect effects, but does not include a time-lagged independent variable. It therefore makes no distinction between short-term and long-term effects:

$$Y_t = \rho W Y_t + X_t \beta + W X_t \theta + \alpha + u_t$$

The matrix of partial derivatives is simply:

$$\left[\frac{\partial E(Y)}{\partial x_{1k}} ... \frac{\partial E(Y)}{\partial x_{Nk}} \right] = (I - \rho W)^{-1} (\beta_k I_N + \theta_k W)$$

where the average direct effect is given by the mean diagonal element of the matrix and the average indirect effect is given by the mean row sum of the non-diagonal elements. The total effect, given by the sum of the previous two effects, is the average impact of a regressor X_k on the dependent variable.

Appendix C: Alternative specifications of the spatial weight matrix

Table 12: Comparison of spatial weight matrices - Log-likelihoods

	Spatial weight matrix	**SDM**	**DSDM**
Overall sample	First-order contiguity	2733.61	2774.97
	Second-order contiguity	2722.62	2766.13
	3 nearest neighbours (3NN)	2717.95	2757.18
	4 nearest neighbours (4NN)	2722.81	2765.81
	5 nearest neighbours (5NN)	2728.07	2770.23
	6 nearest neighbours (6NN)	2724.58	2764.72
	7 nearest neighbours (7NN)	2721.44	2759.44
	Inverse distance (max 100km)	2732.21	2776.09
Male sample	First-order contiguity	2577.47	2591.04
	Second-order contiguity	2577.70	2594.22
	3 nearest neighbours (3NN)	2572.79	2586.47
	4 nearest neighbours (4NN)	2575.49	2588.76
	5 nearest neighbours (5NN)	2574.84	2589.23
	6 nearest neighbours (6NN)	2577.26	2592.02
	7 nearest neighbours (7NN)	2582.63	2598.10
	Inverse distance (max 100km)	2584.94	2599.76
Female sample	First-order contiguity	2489.72	2533.36
	Second-order contiguity	2468.43	2515.82
	3 nearest neighbours (3NN)	2476.54	2515.35
	4 nearest neighbours (4NN)	2478.77	2521.12
	5 nearest neighbours (5NN)	2487.04	2526.09
	6 nearest neighbours (6NN)	2476.12	2516.68
	7 nearest neighbours (7NN)	2474.57	2515.72
	Inverse distance (max 100km)	2479.61	2522.48
Youth sample	First-order contiguity	2149.23	2164.84
	Second-order contiguity	2149.87	2165.22
	3 nearest neighbours (3NN)	2146.09	2159.71
	4 nearest neighbours (4NN)	2145.73	2161.61
	5 nearest neighbours (5NN)	2143.65	2158.95
	6 nearest neighbours (6NN)	2142.67	2159.26
	7 nearest neighbours (7NN)	2141.08	2158.34
	Inverse distance (max 100km)	2149.03	2164.43

Table 13: Direct, indirect and total effects for SDM - Overall sample

		Long-term effects		
	Variables	Direct	Indirect	Total
Inv. distance	% of 25-34 year-olds in the labour force	0.008	0.262***	0.270***
	% of workers living outside their area of work	-0.016	-0.051	-0.067
	Number of microfirms per 100000 inhabitants (log)	0.020***	0.017*	0.037***
	Number of newly established firms per 100000 inhabitants (log)	-0.011***	-0.036***	-0.047***
	% of employment in the agricultural sector	-0.045	-0.706***	-0.751***
	% of employment in the industrial sector	-0.009	-0.390***	-0.400***
	% of employment in the construction sector	-0.061	-0.258	-0.319
	% of employment in the sales sector	0.001	-0.413***	-0.412***
	% of employment in the education, health and public sector	-0.112***	-0.154	-0.266**
	Excess supply of tertiary graduates	0.105***	-0.038	0.067**
	% of temporary contracts	0.009	-0.301***	-0.292***
	Unemployment rate	-0.031	0.172	0.141
5NN	% of 25-34 year-olds in the labour force	-0.005	0.133**	0.127*
	% of workers living outside their area of work	0.009	0.186***	0.196***
	Number of microfirms perr 100000 inhabitants (log)	0.028***	0.002	0.030***
	Number of newly established firms per 100000 inhabitants (log)	-0.010***	-0.005	-0.015**
	% of employment in the agricultural sector	-0.015	0.009	-0.007
	% of employment in the industrial sector	0.024	0.003	0.027
	% of employment in the construction sector	-0.066	-0.129	-0.194
	% of employment in the sales sector	0.001	-0.197***	-0.196**
	% of employment in the education, health and public sector	-0.092***	0.023	-0.070
	Excess supply of tertiary graduates	0.107***	-0.027	0.080***
	% of temporary contracts	0.037	-0.077	-0.041
	Unemployment rate	-0.046	0.240***	0.194**

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

Table 14: Direct, indirect and total effects for DSDM - Overall sample

		Short-term effects			**Long-term effects**		
	Variables	Direct	Indirect	Total	Direct	Indirect	Total
Inv. distance	% of 25-34 year-olds in the labour force	0.017	0.293***	0.310***	0.025	0.405***	0.430***
	% of workers living outside their area of work	-0.016	-0.043	-0.058	-0.016	-0.066	-0.082
	Number of microfirms per 100000 inhabitants (log)	0.018***	0.014	0.032***	0.017***	0.027**	0.044***
	Number of newly established firms per 100000 inhabitants (log)	-0.011***	-0.036***	-0.047***	-0.012***	-0.053***	-0.065***
	% of employment in the agricultural sector	-0.048	-0.733***	-0.781***	-0.068	-1.016***	-1.083***
	% of employment in the industrial sector	-0.017	-0.424***	-0.442***	-0.029	-0.584***	-0.613***
	% of employment in the construction sector	-0.062	-0.313	-0.376*	-0.069	-0.452*	-0.520*
	% of employment in the sales sector	-0.000	-0.385***	-0.385***	-0.012	-0.522***	-0.534***
	% of employment in the education, health and public sector	-0.119***	-0.155	-0.274***	-0.118***	-0.261*	-0.379**
	Excess supply of tertiary graduates	0.103***	-0.078**	0.024	0.095***	-0.062	0.033
	% of temporary contracts	0.012	-0.330***	-0.318***	0.001	-0.443***	-0.442***
	Unemployment rate	-0.028	0.102	0.074	-0.024	0.125	0.101
5NN	% of 25-34 year-olds in the labour force	0.003	0.129**	0.132*	0.011	0.183**	0.195*
	% of workers living outside their area of work	0.004	0.205***	0.209***	0.017	0.291***	0.308***
	Number of microfirms per 100000 inhabitants (log)	0.024***	-0.002	0.021***	0.023***	0.008	0.031***
	Number of newly established firms per 100000 inhabitants (log)	-0.009***	-0.001	-0.010	-0.009***	-0.005	-0.014
	% of employment in the agricultural sector	-0.017	0.044	0.027	-0.014	0.054	0.040
	% of employment in the industrial sector	0.014	-0.009	0.005	0.014	-0.006	0.007
	% of employment in the construction sector	-0.075	-0.151	-0.226*	-0.084*	-0.249	-0.333*
	% of employment in the sales sector	0.006	-0.157**	-0.151**	-0.004	-0.218**	-0.222**
	% of employment in the education, health and public sector	-0.105***	0.038	-0.067	-0.102***	0.003	-0.099
	Excess supply of tertiary graduates	0.102***	-0.057***	0.045**	0.097***	-0.032	0.065**
	% of temporary contracts	0.033	-0.096	-0.063	0.027	-0.119	-0.092
	Unemployment rate	-0.037	0.169*	0.131	-0.026	0.219*	0.193

*** indicates significance at the 1% level, ** at the 5% level and * at the 10% level.
Source : Author's calculation based on French population census and REE

Appendix D: Cartography of French employment areas

Figure 4: French employment areas - 2010 zoning