Document de travail

The spatial structure of French wages: Investigating the robustness of two-stage least squares estimations of spatial autoregressive models

N° 2008-03
Février 2008

Sylvain Barde
OFCE-DRIC

Observatoire Français des Conjonctures Économiques
69, Quai d'Orsay 75340 Paris Cedex 07
Tel : 01 44 18 54 00 Fax : 01 45 56 06 15
E-mail: ofce@ofce.sciences-po.fr Web: http://www.ofce.sciences-po.fr

The spatial structure of French wages:
Investigating the robustness of two-stage least squares estimations of spatial autoregressive models

Sylvain Barde
OFCE-DRIC
250 rue Albert Einstein
06560 Valbonne
sylvain.barde@ofce.sciences-po.fr

Abstract :

Two stage least squares are a popular method of estimation of spatial auto-regressive models, where the dependent variable in an area is a function of the value of the same variable in contiguous areas. Existing literature on this topic points out, however, that this creates problems of consistency. Nevertheless, studies such as Fingleton (2003) show that such an approach is being used to test the central hypothesis of New Economic Geography that increasing returns to agglomeration lead to the concentration of economic activity. It is therefore important to investigate the validity of the methodology in this case.

The focus of this study is twofold: first to replicate the methodology of Fingleton (2003) on the French case and investigate the presence of increasing returns to agglomeration in the spatial structure of wages in France. Secondly, because of the econometric problems inherent to the specification pointed out in the literature, the study tests the validity and robustness of the results obtained.

The first central finding is the significant presence of such returns to scale for France, similar to the ones found in the UK and in other studies of French spatial wage disparities. The second finding is that rigorous tests on the instrumentation strategy defined in Fingleton (2003) reveal that the instruments are typically strong and lead to consistent estimates. Finally, the methodology is shown to be robust to changes in the specification of the spatial weights matrix, and that taking into account a larger time-dimension through a simple pooled regression is valid and leads to an improvement of the significance of the parameters.

Keywords : *spatial econometrics, increasing returns, spatial autoregressive model*
JEL Codes : *C21, R12, R23*

The spatial structure of French wages:
Investigating the robustness of two-stage least squares estimations of spatial autoregressive models

1. Introduction

Increasing returns to agglomeration are a fundamental concept of the literature on the localisation of economic activity. Given the presence of competitors, of congestion costs or negative externalities, the existence of concentrated of economic activity is only possible if agents receive a benefit from such agglomeration. For Fujita and Thisse (1996), increasing returns to agglomeration are essential in explaining the geographic distribution of economic activity. Fujita et al (1999) take this statement to be the foundation of New Economic Geography (NEG). An economy functioning with constant returns to scale on its geographical dimension would be dispersed over space; in a situation they call “backyard capitalism”.

Although a lot of emphasis is put on such increasing returns to agglomeration as an explanation for the observed concentration of economic activity, only recently have empirical studies attempted to quantify them. While Fujita and Thisse (1996), Krugman (1998) or Ottaviano and Puga (1999) call for more empirical testing of the theoretical predictions, Head and Mayer (2004) or Brakman et al (2006) start reviewing the results of such empirical studies. Amongst this recent group of studies that test for the presence of increasing returns to scale, Fingleton (2003) stands out as using a model specification that accounts for both increasing returns to scale and spatial spillovers, whilst using a straightforward econometric approach, two-stage least squares (2SLS), and easily available data. The central conclusion of this study is that it is possible to isolate a significant level of increasing returns to employment density, and therefore agglomeration, in the United Kingdom.

However, given that this econometric model allows for spillovers, the wages in an area depend on the wages in neighbouring areas. This means that the specification used in the estimation is in fact a spatial auto regressive (SAR) model. Kelejian and Prucha (1997, p103-104) show that while the use of 2SLS in such cases is "suggestive, computationally convenient and therefore tempting", the results are not consistent in general. Kelejian and Prucha (1998) do however provide suggestions to get around this problem, particularly with respect to the choice of instruments, and part of this advice is integrated into the 2SLS estimation strategy of Fingleton (2003).[1] Indeed, according to Kelejian and Prucha (1998), for all its econometric shortcomings, the 2SLS-based estimation of SAR models is computationally and intuitively simple, and feasible for large samples, while maximum likelihood, for example, is not. It is therefore important to establish the robustness of such estimations.

The first aim of this study is therefore to use the French case to replicate the methodology developed in Fingleton (2003), and compare the estimates both with the original study and more recent work on France. Once this is done, the second aim is to test the robustness of this methodology, by checking the validity of the instrument set and evaluating the sensitivity of the estimations to a change in the spatial and temporal dimensions of the problem.

The remainder of the paper is organised as follows: Section 2 briefly reviews the econometric model and methodology used in Fingleton (2003), and section 3 presents the initial results obtained for France following this methodology. Section 4 discusses the validity of the instruments, and section 5 then presents the robustness tests on the spatial weight

[1] Kelejian and Prucha (1998) show in particular that if the instrument set contains at least first order spatial lags of the exogenous variables, then the 2SLS estimator is consistent. While they also show that this estimator does not use the extra information available in the spatial auto-correlation of the error term, this is sufficient to correct the inconsistency problem of the 2SLS estimator.

matrix. Section 6 shows how the results are affected when a larger temporal dimension is introduced. Finally, section 7 concludes.

2. The Fingleton (2003) methodology: a 2SLS estimation of a SAR wage equation

The spatial literature mentioned in the introduction focuses on the central role played by heterogeneity-based increasing returns to agglomeration in explaining the agglomeration mechanisms and spatial structures of the economy, often through the simplification of a Dixit-Stiglitz (1997) model of monopolistic competition. In order to generate such effects, Fingleton (2003) uses a model of final production of goods and services Q with inputs M, labour in the final production sector measured in efficiency units, and I, the intermediate consumption of a CES composite of producer services.

$$Q = \left(M^{\beta} I^{1-\beta}\right)^{\sigma} \tag{1}$$

This equation can be re-written to take into account the whole production process by aggregating the final and intermediate productions:

$$Q = \phi N^{\gamma} \tag{2}$$

Where $N = EA$ is total labour in efficiency units and $\gamma = \sigma\left(1 + (1-\beta)(\mu - 1)\right)$. E is the employment density in a given area, and serves as a measure of the concentration of economic activity, and A is the efficiency of labour. The elasticity γ gives the overall level of returns to density. In particular, within the γ term, μ represents the existence of increasing returns to agglomeration in the production of the producer services I. [2] In this model, the increasing returns to agglomeration stem from the increase in the variety of producer services with employment density, through the preference for variety that exists in the CES composite I.

[2] For a more complete derivation of the econometric model from the initial theoretical model, the reader is referred to the appendix of Fingleton (2003)

The motivation behind the use of wages as the main dependant variable of spatial econometrics is that there is little data on value added at a disaggregated spatial level, whereas it is possible to find wage data at these same spatially disaggregated levels. This implies that it is difficult to estimate directly a linearised version of equation (2), and that it is preferable to differentiate it with respect to *N* to obtain the wage equation. Differentiating, taking logs and substituting the original definitions of *Q* and *N* gives the following wage equation:

$$\ln(w) = \alpha + (\gamma - 1)\ln(E) - (\gamma - 1)\ln(A) \qquad (3)$$

Where the constant is $\alpha = \ln(\gamma) + \ln(\phi)$. For the moment, the specification of the model still depends on *A*, the efficiency of labour in each area. This is not directly observable, however, and must therefore be proxied for. To do so, Fingleton (2003) includes two more variables, which are *H*, the level of education in each area, and *T*, an index of the technological intensity of the area. Furthermore, because workers are mobile between areas and the wage variable comes from firm data, the possibility of spillovers between regions needs to be taken into account. Corrected for these elements, the initial specification of the econometric model used in Fingleton (2003) is:

$$\ln(w) = \alpha + \beta W \ln(w) + (\gamma - 1)\big(\ln(E) - \beta W \ln(E)\big) + \delta_1 H + \delta_2 T + \varepsilon \qquad (4)$$

One of the central advantages of this SAR specification is that the estimated parameter for increasing returns to agglomeration is (γ-*1*) and not γ. In the absence of returns to scale, γ is equal to one, and the composite employment density variable $\ln(E) - \beta W \ln(E)$ disappears from the equation. This allows for a direct test of the presence of such returns to scale, simply by looking at the sign and statistical significance of the parameter on the composite employment density variable, which measures the effective concentration of agents in space.

As mentioned in the introduction, the econometric methodology used to estimate this SAR wage equation is an iterated two-stage least squares (2SLS) estimation. This is required for two reasons. First of all, the independent variable, the log of wages w, and the dependant variable E, employment density, are likely to be jointly determined, which means that one of regressors is an endogenous variable. Secondly, the literature, such as Anselin (1988) and Fingleton (1999), shows that ordinary least squares (OLS) are inconsistent when a spatially lagged dependent variable is included amongst the regressors. This is the case here with the variable $W\ln(w)$, which is meant to capture the spatial spillover of wages. Both these reasons justify the use of 2SLS in Fingleton (2003).

The use of this procedure, however, means that instruments need to be provided for the two variables $W\ln(w)$ and $\ln(E)$. The method used in Fingleton (2003), and replicated here is the triple group method of Kennedy (1992). This involves the creation of an instrument, named E_I which takes the value -1, 0 or 1 for an employment area depending on whether the value of E is ranked in the lowest, middle or highest third. This approach allows the creation of instruments that do not require extra variables or information beyond the one already available. Following Fingleton (2003), the other instruments included are the two exogenous variables H and T, as well as the spatial lags of these three instruments, WE_I, WH and WT. These last two variables are specifically included to instrument the spatial lag of the dependent variable, $W\ln(w)$. As mentioned in the introduction, Kelejian and Prucha (1998, p107) show that a 2SLS estimation of a spatial autoregressive model will be consistent if the instrument set includes at least first order spatial lags of all the exogenous variables. This is why Fingleton (2003) uses WH and WT, and why they are also included here. Part of the robustness checks carried out later in this paper involve determining the quality of these instruments in the estimations carried out.

A final methodological consideration is the iteration procedure used as part of the estimation. In order to estimate correctly the returns to scale parameter (γ-1), the spillover parameter β on the spatial lag of employment $W\ln(E)$ in equation (4) must be equal to the spillover parameter on the spatial lag of wages $W\ln(w)$. This is achieved by iterating the estimation. The starting value of the composite employment density variable is arbitrarily set to be $\left(\ln\left(E\right)-W\ln\left(E\right)\right)$, and initial estimate of the spillover β obtained is used to recalculate the density variable. Further estimations refine the estimate of β and this iterative process continues until the incremental improvements in β fall below a given tolerance, set here at 10^{e}-4.

Before presenting the results of these estimations, it is important to discuss some elements relating to the French data used in the following estimations. The geographical units used in this paper are the French employment areas, which divide the surface of mainland France into 348 units, and in order to carry out the robustness analysis mentioned in introduction, five years of data were collected (2000-2004), compared to the two years used in Fingleton (2003). The specific sources for the wage, employment, education and technology variables are detailed in Appendix 1. Furthermore, an illustration of the potential relationships between these variables can be seen in the scatter provided in appendix 2.

It is important to point out that the database used to determine the average wage per employment area contains firm-level data and not plant-level data. This means that for multi-plant firms, the geographical localisation of all the plants is that of the firm headquarters. Because the headquarters of multi-plant firms are mainly based in urban areas, including such firms would over-emphasise the importance of urban areas. In order to avoid this problem, only single plant firm data are used in the calculation of the average wage data. However, if on average multi-plant firms pay higher wages than single plant firms, this will under-estimate the size of wages in the employment areas where these plants are based. The potential biases induced by this measurement problem will be discussed in the next section.

It is also necessary to detail how the spatial weights matrix W is obtained. As was explained above, the W matrix reflects the fact that workers do not necessarily reside and work in the same employment area, and hence they commute to work across the borders of the geographical units used in this analysis. The aim of the spatial lags introduced with the help of the *W* matrix is therefore to replicate the spatial connectivity of French employment areas, which serves as the basis of the estimation of the potential spillovers that occur between regions as a result of this commuting to work.

This matrix is defined as follows, using a simple point to point distance matrix and a decay rate τ :

$$W_{i,j} = \exp\left(-\tau_i d_{i,j}\right) \quad \text{for } i \neq j$$

$$W_{i,j} = 0 \quad \text{for } i = j \text{ or } d_{i,j} > 100\text{km}$$

The diagonal elements of the *W* matrix are set to zero, as well as those cells representing distances above 100km, so that only the potential influence of relatively close regions is taken into account. Furthermore, as is the case in Fingleton (2003), the decay rate τ_i that weighs the elements of the spatial matrix is specific to each employment area, and is calibrated with commuting data. For each employment area i, the decay rate is calibrated using an iterative process to minimise the following relation:

$$\sum_j \left(P_j - N_j \exp\left(-\tau_i \times d_j\right)\right)^2$$

Here, d_j is a given distance band[3], P_j is the proportion of French workers that commute this distance on average every day to go to work, and N_j is the number of neighbouring employment areas included within the distance band. For each employment area i, the decay rate τ_i is the one that reproduces best the local structure of commuting. It is important to point out that this calibration is less detailed than the one done in Fingleton (2003). Indeed, in the original study the parameter τ_i for each geographical unit is calibrated with commuting data that is specific to that unit, whereas here it is the French national average that is used. The

[3] There is commuting data for 16 bands of 5km, covering a radius of 0-80 km and one additional band for « more than 80km ».

robustness test in section 5 will show, nevertheless, that this problem has little impact on the results of the estimation.

3. Preliminary results

Initial estimations were carried out using the methodology and data outlined above, in order to generate a set of benchmark estimates for the French case. Table 1 shows the results of the estimation of equation (4), treating each year of observations as a separate cross-section. Apart from the constant term, the only significant variable for all years is the returns to scale parameter. The education and technology variables only seem significant for some years, and the spatial spillovers term do not seem to contribute anything significant to the estimation. Furthermore, diagnostic tests carried out on these estimations reveal some problems. A first test reveals the presence of a strong correlation between the dependant variable ln(w) and the residuals, suggesting the existence of omitted variables. In order to test for the presence of spatial auto-correlation in the residuals of the estimation, Moran's I test was carried out on the residuals of the estimations. The specification used for the distribution of the statistic is the one developed in Anselin and Kelejian (1997), as the presence of an endogenous regressor and a spatial lag needs to be taken into account.[4] The tests carried out here suggest there is no significant spatial auto-correlation in the residuals.

These diagnostics are in line with the results reported in Fingleton (2003), although his results show less correlation between wages and residuals, and more spatial autocorrelation.[5] The interpretation of the high correlation between wages and residuals made in the Fingleton

[4] Anselin and Kelejian (1997) show that under these conditions, the standard moments method of Cliff and Ord (1973), which uses OLS and not an IV method, gives inconsistent results.
[5] Fingleton 2003 reports a correlation between wages and residuals of 0.5279 for 1999 and 0.5495 for 2000. The Moran statistics are 2.689 for 1999 and 1.002 for 2000, which suggests the existence of spatial autocorrelation for 1999, but not for 2000.

study is that there are unobservable variables that affect the productivity of labour. These are in particular knowledge spillovers and spatial externalities *à la* Jacobs/Marshall, which are unobservable and therefore omitted. In order to correct this problem, Fingleton suggests introducing the lagged residuals of model (4) into the estimation :

$$\ln(w) = \alpha + \beta W \ln(w) + (\gamma - 1)\left(\ln(E) - \beta W \ln(E)\right) + \delta_1 H + \delta_2 T + \delta_3 r_{-1} + \varepsilon \qquad (5)$$

Table 1
Preliminary cross-sectional regression, dependant variable : log of wages

Variable	Parameter	2000	2001	2002	2003	2004
Constant	α	5.8144***	5.8290***	5.8619***	5.8728***	5.9368***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Spatial spillover	β	0.0085	0.0130	0.0039	0.0091	0.0024
		(0.3158)	(0.1096)	(0.6516)	(0.2531)	(0.7469)
Returns to density	γ-1	0.0438***	0.0413***	0.0448***	0.0319***	0.0324***
		(0.0000)	(0.0000)	(0.0000)	(0.0003)	(0.0001)
Education	δ_1	0.0354	0.0899	0.0513	0.2865**	0.1587
		(0.7746)	(0.4585)	(0.6936)	(0.0178)	(0.1644)
Technology	δ_2	0.0143	0.0139	0.0267***	0.0131	0.0149*
		(0.1053)	(0.1069)	(0.0039)	(0.1242)	(0.0627)
Diagnostics						
N° of observations		348	348	348	348	348
R^2		0.2563	0.2805	0.2770	0.2800	0.2293
Correlation ln(w) - residuals		0.8338	0.8180	0.8206	0.8323	0.8528
Standardised Moran's I		0.7463	0.9073	0.6573	0.6220	0.8749
Prob (H0 = no spatial auto-correlation)		(0.4555)	(0.3643)	(0.5110)	(0.5340)	(0.3817)

The lagged residuals r_{-1} are used here as an instrument measuring the realisation of the externalities omitted from the initial specification. The estimation methodology remains unchanged, apart from the fact that these residuals are likely to be themselves correlated with the error term, and therefore need to be instrumented. The extra instruments are generated using the same methodology as above: $(r_{-1})_I$ takes value -1, 0 or 1 depending on the ranking of r_{-1}, and $W(r_{-1})_I$ is the spatial lag of this instrument.

As for the previous specification, this extension is estimated as a cross-section, using for each year the lagged residuals of the Table 1 estimations as a measure of omitted variables. Compared to the results found in Table 1, the estimates in Table 2 do not change much, but turn out to be statistically more significant. The parameter on the lagged residuals, the "omitted variables", is itself very significant which confirms that the initial theoretical model is not enough to explain the spatial variations of wages. As far as the diagnostic tests are concerned, although the correlation between the log of wages and the residuals is still present, it is much lower than in the first round of estimations.

Table 2
Cross-sectional regression with residuals, dependant variable : log of wages

Variable	Parameter	2001	2002	2003	2004
Constant	α	5.8287***	5.8622***	5.8729***	5.9370***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
Spatial spillover	β	0.0133***	0.0051	0.0094*	0.0033
		(0.0026)	(0.2682)	(0.0730)	(0.4478)
Returns to density	γ-1	0.0412***	0.0430***	0.0314***	0.0310***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
Education	δ_1	0.0902	0.0658	0.2907***	0.1702**
		(0.1679)	(0.3387)	(0.0003)	(0.0101)
Technology	δ_2	0.0139***	0.0270***	0.0133**	0.0152***
		(0.0030)	(0.0000)	(0.0183)	(0.0012)
Omitted variables	δ_3	0.7783***	0.8713***	0.7040***	0.7750***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
Diagnostics					
N° of observations		348	348	348	348
^{2}		0.7907	0.7983	0.6921	0.7428
Correlation ln(w) - residuals		0.5004	0.4918	0.5505	0.4864
Standardised Moran's I		0.20935	-0.0821	-0.3541	-0.0533
Prob (H0 = no spatial auto-correlation)		(0.8342)	(0.9346)	(0.7233)	(0.9575)

This improvement in the diagnostics and the statistical significance of the estimated parameters following the introduction of the lagged residuals is also reported in Fingleton (2003). Furthermore, the value of the parameters found here is on the same order as those reported by Fingleton.[6] This suggests that the results of the original study have been reproduced here, although the overall significance is lower.

The main difference with the results of the Fingleton (2003) estimations on the UK is that the spatial spillovers, measured by the β parameter, are less significant. The first possible explanation to this is the measurement bias mentioned in the previous section. By using firm level data rather than plant level data, the measure of average wages will be biased downwards in those employment areas that house the plants of multi-plant firms, in great majority urban areas. This is somewhat visible in the scatter plots in appendix, which appear to be flatter than the ones provided in the Fingleton study.

The second explanation behind the relative lack of significance of the spatial spillovers is the lower resolution of the *W* matrix. The Fingleton (2003) analysis uses 408 local authorities for a country, the UK, which has a surface of 245,000 km^2. Our data is based on 348 employment areas over 550 000 km^2. The measured importance of spillovers depends on the number of people that cross a border between two geographical units when they commute to work. This number will of course be larger the finer the spatial grid used, and two given locations with a given commuting to work pattern will thus appear closer the finer the grid. Such a change in the *W* matrix might have an influence on the estimation of the spillovers. In a paper on the concentration of economic activity, Duranton and Overman (2005) show that the use of discrete spatial data introduces a bias the size of which depends on the resolution of the grid. In their work they manage to avoid this problem by using concentration indexes that use a continuous spatial dimension rather than a discrete one. Unfortunately, in our case, the very definition of the *W* matrix depends on the existence of a finite number of locations. It is partly to evaluate the importance of this bias that robustness tests are carried out on the spatial weight matrix *W* in section 5.

---

6 The parameters reported in Fingleton (2003) are α =5.5460 ; β =0.0014 ; $(\gamma\text{-}1)$ = 0.016 ; δ_1 = 0.2929; δ_2 = 0.0503 and δ_3 = 0.7762.

The results in Table 2 suggest the significant presence of increasing returns to agglomeration, with a γ parameter equal to 1.03-1.04. This implies that having controlled for the level of education, technological intensity, and the existence of commuting, the differences in wages, and therefore implicitly in labour productivity, are more than proportional to the differences in density. Furthermore, as was pointed out in section 2, the γ parameter is an elasticity, which means that the premium on density is 3 - 4% of the variation in density. This estimate is in line with recent research by Combes *et al* (forthcoming), who find a similarly sized elasticity of wages with respect to the density of employment for France using a sorting approach on a large panel of French workers. The consistency of the estimates found here with both the original Fingleton study and these recent estimates for France of Combes *et al* suggests that the 2SLS estimation of the SAR wage equation using the instrumentation strategy of Fingleton (2003) produces consistent results.

4. Validity of the instruments

At this point, the focus of the paper moves beyond the replication of the Fingleton (2003) methodology towards checking robustness of the methodology itself. As mentioned previously, the first of these checks relates to the validity of the instruments, as this is a critical element of the 2SLS estimation. The reliability of IV estimators in general rests on two central hypotheses. The first is that the instruments are exogenous, and therefore uncorrelated with the error term. The fact that one of the main instruments, E_I, is a simple transformation of the employment density variable $\ln(E)$, which is itself assumed to be correlated with the error term, raises the question of a possible violation of this hypothesis. The same is true of $(r_{-1})_I$ with respect to r_{-1}. The second hypothesis underpinning the 2SLS estimator is that the instruments have to be relevant. This is an essential aspect, as Staiger and Stock (1997) show that 2SLS estimations are not consistent if the instruments are weak, in other words if the relevance of the instruments with respect to the endogenous regressors falls below a certain critical level.

The importance of testing the instruments goes beyond purely econometric requirements for consistent estimation. Indeed, the estimation strategy rests on the use of instruments that are transformations of the instrumented variables. The gist of this strategy is to instrument endogenous regressors with a triple group transformation, following Kennedy (1992), and spatial lags of the dependant variable with spatial lags of the exogenous variables, following the recommendations of Kelejian and Prucha (1998). Hence, the key benefit is that no external information needs to be provided. If either hypothesis turns out to be violated, the econometric consequence is that the 2SLS estimates will be biased. In that case, however, the deeper implication is that the whole instrumentation strategy needs to be re-thinked.

First of all, in order to test the hypothesis of exogeneity, a Sargan test is carried out in all the regression carried out in this study. Two approaches are then used to test the requirement of relevant instruments. The first is the rule of thumb developed by Staiger and Stock (1997), and the second is the more formal test of Stock and Yogo (2005). The rule of thumb states that a set of instruments is considered to be weak if the first stage F statistic is less than 10. Although this informal test is simple and intuitive, it must be carried out separately for all the endogenous regressors that are instrumented. The more formal version of this test, specified in Stock and Yogo (2005), provides a single test statistic for all the endogenous regressors. If there is only one endogenous regressor, the Stock and Yogo (2005) statistic effectively boils down to a first stage F statistic, although Stock and Yogo (2005) also provide critical values that prove more detailed than the simple level of the rule of thumb.

The results of these tests are visible in Table 3. The Sargan tests suggest that for 2001-2003 there is no auto-correlation between the instruments and the error term, even though the instruments E_I and $(r_{-1})_I$ are transformations of *a priori* endogenous variables. For 2004, however, the null is rejected, and the instruments do not seem to be exogenous. This seems to indicate that the use of transforms of existing variables as instruments can indeed create some problems with exogeneity. However, comparing the estimations that pass the Sargan test with those that fail it suggests that the results are not affected, which minimises the impact of this sporadic lack of exogeneity.[7] Nevertheless, the tests do reveal that a potential weakness of this approach to instrumentation is a lack of exogeneity.

Table 3
Instrument tests on the corrected regression (5)

	2001	2002	2003	2004
Exogeneity				
Sargan test (χ^2, 6 df)	3.9218	8.8858	1.5045	14.7005
Prob (H0 = no correlation with error)	(0.6873)	(0.1801)	(0.9592)	(0.0227)
Relevance				
Stock-Yogo statistic (5% critical value = 12.20)	17947.63	17284.00	15119.16	16739.78
First stage F-statistic on $W\ln(w)$	598.40	601.51	602.21	586.38
First stage F-statistic on $\ln(E)$	159.59	163.14	161.46	156.73
First stage F-statistic on r_{-1}	55.67	53.02	46.24	50.88

As far as the relevance of the instruments is concerned, the null assumption of a weak instrument set is systematically rejected. The high value of the Stock-Yogo test statistic can be explained by the fact that the instrument set seems strong for each of the instrumented variables individually. Indeed, the first stage F-statistics of the instrument set are significantly higher than the rule of thumb value of 10, suggested by the Staiger and Stock (1997), under

[7] This is particularly visible further on in Table 6, where one model specification passes the test and the other one fails. The parameter estimates and diagnostic tests, however, remain very close.

which one must doubt of the relevance of the instrument set.[8] Because the Stock and Yogo (2005) test measures the simultaneous relevance of the instrument set on all the instrumented variables ($W\ln(w)$, $\ln(E)$ and r_{-1} in this case), it is normal that the test statistic be larger. Intuitively, the probability that the instruments are weak for all three instrumented variables is multiplicatively smaller than the probability that the set is weak on a single one of them.

The strong relevance of the instruments suggested by these tests probably stems from the very fact that they are transforms of existing variables of the model, either by using the three group method of Kennedy (1992) or through a spatial lag, such as the *WH* and *WT* variables. In this respect, although the instrumentation strategy mentioned previously might create problems with exogeneity, it generates by construction a strong set of instruments. As one can see from the F-tests, this is particularly true for $W\ln(w)$, which confirms the recommendation of Kelejian and Prucha (1998) that spatial lags of the exogenous variables be included in the regression to instrument the spatial lag of the dependant variable.

5. Spatial weights and robustness of the results

The next the robustness tests carried out is a change is the calibration of the spatial weight matrix W. On top of the possible measurement bias due to the resolution of the spatial grid mentioned in section 2, the motivation behind this is the fact that commuting data is not always available depending on the country or region of interest. It is therefore important to establish the sensitivity of the results to the calibration procedure used to determine W. The way this is achieved is to replace the area-specific decay rate τ_i used previously by a unique decay rate τ for all regions. The extended model (5), including the lagged residuals as

[8] The results of the fist stage F-statistic are not reported in the following tables, as they are very regular. Most are of the same size of the ones presented above, and therefore still satisfy Staiger and Stock's rule of thumb.

measures of omitted variables is re-estimated, but with a W matrix calculated with a common decay rate of 0.025.

Table 4
Cross-sectional regression with residuals, τ = 0.025, dependant variable : log of wages

Variable	Parameter	2001	2002	2003	2004
Constant	α	5.8344***	5.8623***	5.8771***	5.9380***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
Spatial spillover	β	0.0012***	0.0007	0.0006	0.0004
		(0.0040)	(0.1236)	(0.2166)	(0.3191)
Returns to density	γ-1	0.0439***	0.0443***	0.0353***	0.0311***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
Education	δ_1	0.0667	0.0498	0.2615***	0.1667**
		(0.3087)	(0.4703)	(0.0010)	(0.0116)
Technology	δ_2	0.0110**	0.0248***	0.0117**	0.0141***
		(0.0214)	(0.0000)	(0.0418)	(0.0033)
Omitted variables	δ_3	0.7731***	0.8636***	0.7029***	0.7714***
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
Diagnostics					
N° of observations		348	348	348	348
R^2		0.7904	0.7971	0.6914	0.7433
Correlation ln(w) – residuals		0.5037	0.4980	0.5508	0.4934
Standardised Moran's I		0.3707	-0.0793	-0.3344	0.0547
Prob (H0 = no correlation with error)		(0.7248)	(0.4863)	(0.9727)	(0.0236)
Sargan test (χ^2, 6 df)		3.6436	5.4603	1.2809	14.603
Prob (H0 = no spatial auto-correlation)		(0.7109)	(0.9368)	(0.7381)	(0.9564)
Stock-Yogo statistic (5% critical value = 12.20)		17725.02	16390.94	14583.63	16853.77

Not only is the decay rate now the same for all the employment areas, but the absolute value of the parameter is set at a level well below that of the previous estimations.[9] This uniform and lower level of the decay rate is designed to produce a stronger and more connected spatial structure, as the $W_{i,j}$ elements of the W matrix will be larger than in the

[9] In tables1, 2, 5 and 6 the average of τ_i on the 348 employment areas is 0.0646. The median value is 0.0616. Robustness tests were carried out with a lower value of τ (0.02, 0.01 and 0.005), but are not reported in detail as the results are very similar to the ones shown in table 3.

previous case. This is also an attempt to imitate the conservative choice that could be made by a researcher that would not have access to commuting data between regions.

The results in Table 4 reveal that the estimation is very similar to the ones carried out previously. There is no large shift in the parameter estimates or in their statistical significance. More importantly, the diagnostic tests do not seem to be greatly affected by the change in towards a less elaborate decay rate, and the instruments validity tests stay similar to those of Table 2. This is an important result, as the change in the connectivity of the *W* matrix itself following the change in the decay rate is quite important, and this should affect all the variables and instruments that are spatially lagged. The fact that the parameter estimates and their significance are not really influenced by this change in the weighting of distances tends to minimise the importance of missing or badly measured commuting data.

6. Introducing a larger temporal dimension

The final test of robustness carried out was to introduce a larger temporal dimension into the econometric model. Indeed, all the estimations carried out until now are cross sectional, even the ones that include the residuals from the previous years. In this case, these lagged residuals are supposed to serve as an instrument measuring the realisation of unobservable, and therefore omitted, externalities. However, even though these residuals serve as an instrument, the corrected model (5) implicitly has an auto-regressive distributed lag (ADL) specification:

$$\ln(w_t) = \alpha + \beta W \ln(w_t) + (\gamma - 1)\left(\ln(E_t) - \beta W \ln(E_t)\right) + \delta_1 H_t + \delta_2 H_t + \delta_3 r_{t-1} + \varepsilon_t \quad (6)$$

With :

$$r_{t-1} = \ln(w_{t-1}) - \alpha - \beta W \ln(w_{t-1}) - (\gamma - 1)\left(\ln(E_{t-1}) - \beta W \ln(E_{t-1})\right) - \delta_1 H_{t-1} - \delta_2 H_{t-1} \quad (7)$$

If one only uses two years of data, as is the case in Fingleton (2003) and in Tables 2 and 4, the presence of these lags reduces the estimation down again to a cross sectional analysis. However, if one wishes to use more than two years of data, then using model (5) with a year by year cross-sectional approach is problematic, as for a given year the specification used to generate the residuals is not the same as the one used to generate the parameter estimates. For example, in Table 2 it is not possible to compare the estimates for 2003 and 2004, as the 2004 estimates implicitly assume a "naïve" regression in 2003, which is different from the one that returns the corrected 2003 parameters.

By combining equations (6) and (7), taking into account the fact that H and T are exogenous and equal for all years (therefore $H_t = H_{t-1}$ and $T_t = T_{t-1}$), one can re-write the residual-corrected model (5) of Fingleton (2003) as an explicit ADL.

$$\begin{aligned}\ln(w_t) = \alpha + \delta_3 \ln(w_{t-1}) + \beta_1 W \ln(w_t) + \beta_2 W \ln(w_{t-1}) + (\gamma_1 - 1)\big(\ln(E_t) - \beta_1 W \ln(E_t)\big) \\ + (\gamma_2 - 1)\big(\ln(E_{t-1}) - \beta_2 W \ln(E_{t-1})\big) + \delta_1 H_t + \delta_2 H_t + \varepsilon_t\end{aligned} \quad (8)$$

Apart from the δ_3 parameter, which should remain the same as in the previous estimations, the α, β_1, β_2, γ_1, γ_2, δ_1 and δ_2 parameters of this specification are estimated freely, and are expected to be different from the corrected model estimates in Tables 2 and 3. Because of the presence of spatial lags as well as endogenous regressors, the estimation methodology remains the same 2SLS approach used for the cross sectional estimations. The only change is that two parameters, β_1 and β_2 in front of (E_t) and $W\ln(E_{t-1})$, need to be constrained by the iterative process instead of one previously. Furthermore, on top of the spatial lag of the dependent variable, there is now also a temporal lag of the dependent variable that needs to be instrumented. In the spirit of the methodology used until now, two extra instruments $\ln(w_{t-1})_I$ and $W\ln(w_{t-1})_I$ are created, based on the same triple group division. As a result, all the right hand side variables in (8) are instrumented except H and T, the instruments for each of these variables being a triple group index and its spatial lag.

Table 5
ADL regression 2001-2004, dependant variable : log of wages

Variable	Parameter	Short run	Long run
Constant	α	1.4729***	5.9436
		(0.0000)	
Spatial spillover	β_1	-1.2796**	0.0514
		(0.0220)	
Lagged spatial spillover	β_2	1.2923**	
		(0.0216)	
Returns to density	$\gamma_1 - 1$	0.0004	0.0309
		(0.8637)	
Lagged returns to density	$\gamma_2 - 1$	0.0077*	
		0.0505	
Education	δ_1	0.0622*	0.2510
		(0.0904)	
Technology	δ_2	0.0048*	0.0194
		(0.0842)	
Lagged log of wages	δ_3	0.7522***	-
		(0.0000)	-
Diagnostics			
N° of observations		1392	-
R^2		0.7484	-
Correlation ln(w) – residuals		0.52493	-
Sargan test (χ^2, 6 df)		10.311	-
Prob (H0 = no correlation with error)		(0.1121)	-
Stock-Yogo statistic *		16162.10	-
Moran Tests			
H0 = no spatial auto-correlation		I Statistic	Prob.
Standardised I 2001		0.0882	(0.9297)
Standardised I 2002		0.0912	(0.9274)
Standardised I 2003		-0.0449	(0.9642)
Standardised I 2004		0.0277	(0.9779)

* Stock and Yogo (2005) do not provide critical values for more than 3 endogenous regressors, and there are 5 here. The high value of the statistic, consistent with the other estimations in the paper leads us the conclusion that the null hypothesis of weakness is rejected.

Table 5 shows the parameter estimates of this regression, as well as the results of the diagnostic tests carried out. The parameters of interest here are the long run parameters, calculated by equalising the time indexes of the ADL specification. Overall, these long run parameters are not very different from the ones obtained previously using a purely cross-sectional approach. Only the increasing returns parameter diverges from the previous estimates, the remaining ones staying of the same order. This specification also shows diagnostic tests similar to the ones in Table 2. The residual correlation is similarly sized and the Moran statistic suggests the absence of spatial auto-correlation. The value for the correlation between the log of wages and the residual is still higher than the one reported in Fingleton (2003). [10] Nevertheless, this estimation shows that it is possible to explicitly use the ADL specification (8) over the entire sample and to obtain estimates similar to the year by year analysis of the corrected specification (5).

This ADL specification, however, creates problems of its own. In particular it is not possible, as a general rule, to calculate the long run parameter for the returns to density (γ-1). This is because the composite density variable $\ln(E_t) - \beta_1 W \ln(E_t)$ is not the same in t and t-1 because the spillover parameters β_1 et β_2 are not the same. This means that when the time indices of the ADL are equalised to calculate the long run parameters, it is not possible to simply add up $\ln(E_t) - \beta_1 W \ln(E_t)$ and $\ln(E_{t-1}) - \beta_2 W \ln(E_{t-1})$. For the specific case of the estimation in Table 4, the long run parameter can be calculated because the lagged composite density variable does not come out as being significant. Whilst this is a fortunate occurrence in this case, it will not be true in general.

The problem stems from the fact that some of the parameters have to be constrained during the estimation, using the iterative procedure. This constraint on the spatial lag of employment density $W\ln(E_t)$ cannot be avoided as it is necessary in order to obtain the

[10] The lowest correlation reported in Fingleton (2003) results from the 2000 estimation, corrected with the residuals of the 1999 estimation. Its value is 0.3159.

estimate of the returns to density parameter (*γ-1*), itself the central objective of the estimation. Even though an ADL seems the most logical econometric specification given the presence of lagged residuals in equation (5), it does not seem to be the most practical alternative given the constraints imposed on the variables.

Table 6
Pooled regression 2001-2004, dependant variable : log of wages

Variable	Parameter	Version (1)		Version (2)	
Constant	α	5.8752***		5.8748***	
		(0.0000)		(0.0000)	
Spatial spillover	β	0.0077***		0.0073***	
		(0.0011)		(0.0064)	
Returns to density	γ-1	0.0367***		0.0377***	
		(0.0000)		(0.0000)	
Education	δ_1	0.1533***		0.1453***	
		(0.0000)		(0.0003)	
Technology	δ_2	0.0173***		0.0171***	
		(0.0000)		(0.0000)	
Omitted variables	δ_3	0.7834***		0.7492***	
		(0.0000)		(0.0000)	
Diagnostics					
N° of observations		1392		1392	
R^2		0.7610		0.6942	
Correlation ln(w) - residuals		0.5131		0.5805	
Sargan test (χ^2, 6 df)		17.4955		10.3849	
Prob (H0 = no correlation with error)		(0.0076)		(0.1094)	
Stock-Yogo statistic (5% critical value = 12.20)		235637.39		297969.92	
Moran Tests					
H0 = no spatial auto-correlation		I Statistic	Prob	I Statistic	Prob
Standardised I 2001		0.2113	(0.8327)	0.2094	(0.8341)
Standardised I 2002		-0.0886	(0.9294)	0.2235	(0.8232)
Standardised I 2003		-0.3491	(0.7271)	-0.4266	(0.6697)
Standardised I 2004		-0.0597	(0.9524)	0.0096	(0.9924)

In order to avoid this problem whilst retaining the positive aspect, which is the integration of the temporal dimension, a pooled regression of specification (5) is carried out on the entire sample. As mentioned previously, from a purely theoretical point of view this is not the best choice of specification because it cannot pick up the dynamic aspects potentially present in the temporal dimension, which a dynamic model can. The assumption that is made, however, is that there is little to no variation of the spatial structure of the French economy over the time dimension of the sample (2000-2004), and that effectively there are no gains to be made by choosing a dynamic structure. Under this assumption, pooling the data does not create a misspecification, and means that the cross-sectional simplicity of the Fingleton specification can be maintained whilst including the whole of the sample in a single estimation.

The pooled estimation is carried out as follows: a first estimation of the simple model (4) is carried out on the pooled 2000-2004 period. This provides residuals that are included in the corrected model (5), which is then estimated for the 2001-2005 period. A second version of the pooled regression is carried out using only the 2000 residuals for all years, instead of the pooled 2000-2004 residuals. This second version gives an indication on the assumption of lack of temporal variation in the sample. Indeed, if there is no such variation, the 2000 residuals can instrument the omitted variables for all years, and the results of both versions should be identical. The results of both these versions are available in Table 6.

The central observation is that the values of the pooled estimates do not change much from the ones estimates in Table 2, and are still in line with the ones reported in Fingleton (2003) for the UK. The improvement brought by the higher number of degrees of freedom in the pooled estimation is that they are now highly significant, including the parameters on spillovers, which were not systematically significant in previous estimations. The second positive aspect in terms of robustness is that the results of the diagnostic tests are still in line with the ones carried out on the corrected estimations of Table 2 and the ADL estimations in Table 5.

A last important result is the comparison between the two versions of the estimation, which serves to validate the pooled regression. Indeed, the estimation is robust to the change in residuals, from 2000-2004 to 2000 for all years. Overall, the diagnostic tests on version (2) are not as good as the ones on version (1), which is to be expected given that there is less information in the (r_{-1}) variable to proxy for omitted variables. The parameters, however, remain significant in this second version, and very close to the ones obtained in the first version. This suggests that there is indeed little variation in the spatial structure of wages over the time period of interest, which would validate the pooled regression over a more time-explicit specification such as the ADL presented above. As mentioned in section 4, this also minimises the impact of the failed Sargan tests, as the parameter estimates are similar in the two versions. Given this relative lack of variation over time, the results in Table 6 suggest that if a medium-sized temporal dimension is available, a simple pooled regression is the best way of integrating it into the Fingleton specification.

7. Conclusion

By using the iterated 2SLS methodology described in Fingleton (2003) on the 348 French employment areas, it is possible to obtain estimates similar to the original ones obtained by Fingleton on the UK, and consistent with existing studies of the French spatial wage disparities. In particular, in line with the theoretical predictions mentioned in the introduction, increasing returns to density or agglomeration seem to be a significant variable explaining the spatial structure of French wages. These results are robust to the choice in the weights of the spatial weights matrix W. An interesting conclusion is that choice of the particular value of the decay rate τ has little influence on the results, as long as the value chosen is lower by about an order of magnitude than the actual values obtainable from commuting data.

Furthermore, a significant literature shows that the use of 2SLS in the estimation of SAR models can be problematic, and in particular depends on the choice and validity of the instruments. A central aspect of this study is therefore to test rigorously the validity of the instrumentation strategy used in the Fingleton methodology. The main finding in this respect is that the transformation of existing variables to generate instruments, in particular spatial lags of the exogenous variables, provides a strong instrument set. In addition, while tests show that such instruments are not always exogenous, this does not seem to affect the results of the estimations.

Finally, if several years of data are available and one wishes to integrate this dimension into the estimation, then several options become available. The central model of the Fingleton (2003) paper, which uses lagged residuals to correct for omitted variables, is effectively an implicit ADL. Though it is possible to estimate this specification and obtain results that are consistent with the cross-sectional approach, the problem is that in general one cannot retrieve the long run parameters on returns to density, due to the constraints imposed as part of the iterated estimation procedure. Given the probable lack of variation of the spatial structure over short to medium term time dimensions, our results indicate that improvements to the diagnostics and significance of the parameters can be obtained simply with a pooled regression.

Bibliography

Anselin L, 1988, *Spatial Econometrics: Methods and Models*, Kluwer.

Anselin L, Kelejian H.H, 1997, “Testing for Spatial Error Autocorrelation in the Presence of Endogenous Regressors”, *International Regional Science Review*, Vol. 20, p153-182.

Brakman S, Garretsen H, Schramm M, 2006, “Putting New Economic Geography to the Test: Free-ness of Trade and Agglomeration in the EU Regions” *Regional Science and Urban Economics*, Vol. 36, p613-635.

Cliff A, Ord J.K, 1973, *Spatial Autocorrelation*, Pion.

Dixit A.K, Stiglitz J.E, 1977, “Monopolistic Competition and Optimum Product Diversity”, *American Economic Review*, Vol. 67, N° 3.

Duranton G, Overman H.G, 2005, “Testing for Localisation Using Micro Geographic Data”, *The Review of Economic Studies*, Vol. 72, N° 253, p1077-1106.

Duranton G, Puga D, 2004, “Micro-foundations of Urban Agglomeration Economies” *The Handbook of Regional and Urban Economics*, J.V.Henderson and J-F Thisse eds, Vol. IV.

Ellison G, Glaeser E.L, 1997, “Geographic Concentration in US Manufacturing Industries: A Dartboard Approach”, *The Journal of Political Economy*, Vol. 105, N° 5, p 889-927.

Fingleton B, 2003, “Increasing Returns: Evidence from Local Wage Rates in Great Britain” *Oxford Economic Papers*, Vol. 55.

Fujita M, Krugman P, Venables A, 1999, *The spatial economy; Cities, Regions and International Trade*, MIT press.

Fujita M, Thisse J-F, 1996, “Economics of Agglomeration”, *Journal of the Japanese and International Economies*, Vol. 10, p339-378.

Head K, Mayer T, 2004, “The Empirics of Agglomeration and Trade”, in *The Handbook of Regional and Urban Economics*, J.V.Hendersom and J-F Thisse eds, Vol. IV.

Kelejian H.H, Prucha I.R, 1997, “Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem”, *International Regional Science Review*, Vol.20, p 103-111.

Kelejian H.H, Prucha I.R, 1998, “A generalised least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances”, *Journal of Real Estate Finance and Economics*, Vol.17, p 99-121.

Kennedy P, 1992, *A Guide to Econometrics*, 3rd edition, Blackwell.

Krugman P, 1998, “Space: The final frontier” *Journal of Economic Perspectives*, Vol.12, N°2.

Staiger D., Stock J.H., 1997, “Instrumental variables regression with weak instruments”, *Econometrica*, Vol 65, N° 3, p557-586.

Stock J.H., and Yogo, M., 2005 “Testing for weak instruments in linear IV regression.” In: Stock J.H. and Andrews D.W.K., Eds, *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg,* Cambridge University Press, Cambridge, pp. 80–108.

Appendix 1 : Data sources

- The wage level *w* is calculated using firm data from the DIANE database for each employment area. Only single plant firms are kept, and the two raw variables used at the firm level are the average annual workforce and the annual wage bill, which allows the calculation of an average weekly wage per employment area.
- The employment density variable *E* is the number of employees in each employment area divided by the surface of the area in square kilometres. The level of employment in each employment area is calculated by the French statistical agency, INSEE, based on firm data, and available freely online. The surface of each employment area is calculated by aggregating the surfaces of the French "communes" that compose it, these surfaces being taken from the "Répertoire Géographique des Communes" (RGC), available from the National Geographical Institute (IGN).
- The education variable *H*, is the proportion of the active population in each employment area with an educational attainment equal to or higher than the baccalaureate. This corresponds to the A-level based measure used in Fingleton (2003). The data on the educational attainment of the active population (aged 15 and above) is taken from the French census of 1999.
- The technological index *T* represents the share of the workforce engaged in ITC activities (NAF code 72) and R&D (NAF code 73) relative to the national average and reflects the specialisation of the workforce of an employment area in high-tech industries. This variable is calculated for 1999, using the same source as the wage data. As for wages, only single plant firms are used.
- For the calculation of the *W* matrix, point to point distances between the centroids of the employment areas are calculated using the RGC, already mentioned above.

Appendix 2 : Scatter plot diagrams

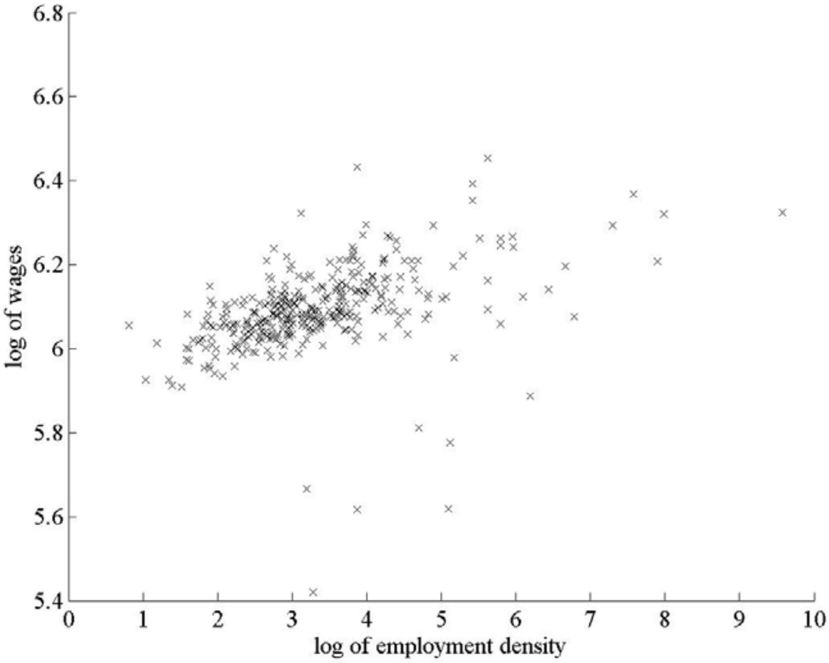

Figure 1 : log of wages vs. log of employment density, 2000

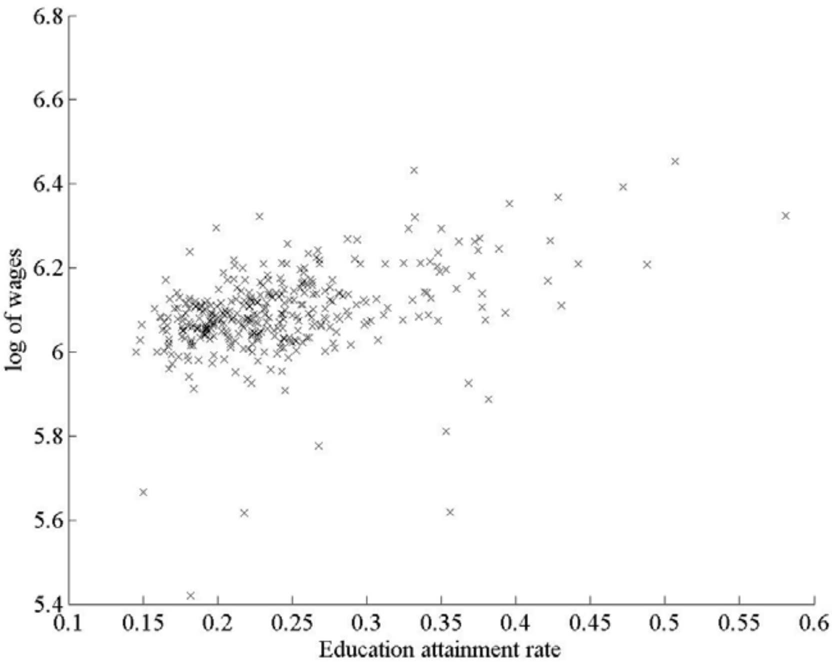

Figure 2 : log of wages vs. educational attainment, 2000

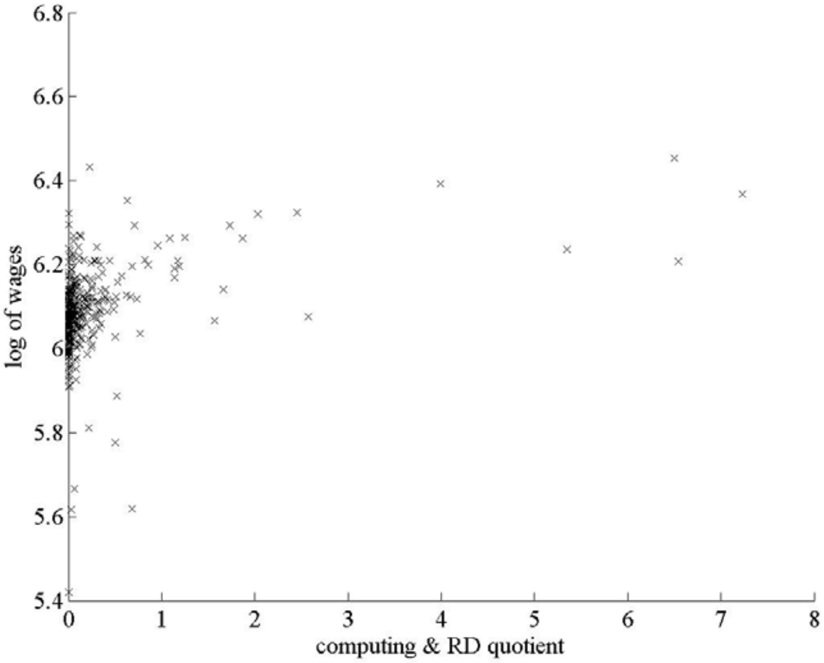

Figure 3 : log of wages vs. IT and R&D quotient, 2000

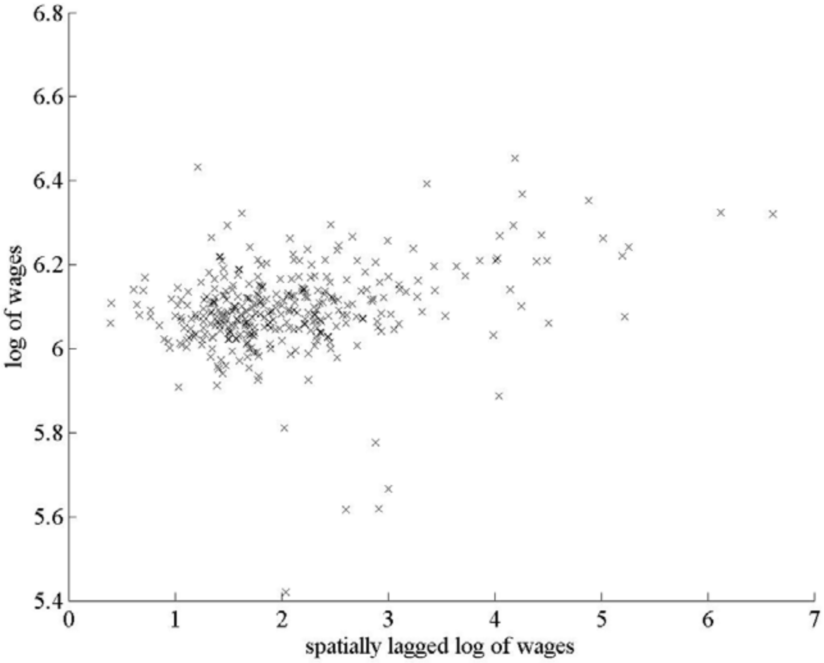

Figure 4 : log of wages vs. spatially lagged log of wages, 2000