The Dynamics of Local Employment in France

Pierre-Philippe Combes* Thierry Magnac† Jean-Marc Robin‡§

First version: July 2002

Revised: April 2003

Abstract

This paper studies the impact of local economic structure on local sectoral employment growth. Local employment growth is decomposed into "internal" growth (the growth of the size of existing plants) and "external" growth (the creation of new plants). Using panel data methods, we estimate the dynamics of both variables simultaneously. Our observations refer to 36 manufacturing, trade and service sectors and 341 French "Employment Areas" in every year between 1984 and 1993.

Carefully specifying the short-run dynamics and controlling for fixed effects and endogeneity are proved to be critical devices. The low order of the selected model, an ARMA(1,1), implies that static externalities are prevalent compared to dynamic ones. Moreover, whereas fixed effects explain most of the spatial variation of plant size, plant creation is mainly determined by the current local economic structure. Policies targeted on the creation of new plants should therefore prove to be more rapidely efficient. For instance, large areas endowed with a small number of even size sectors and where are located large leader firms impulsing growth to smaller and numerous plants experience larger growth.

*CERAS and Boston University. CNRS researcher also affiliated with the CEPR. Boston University, Economics Department - 270, Bay State Road - Boston, 02215 MA, USA. combes@enpc.fr. http://www.enpc.fr/ceras/combes/.

†LEA-INRA and CREST-INSEE. Also affiliated with the CEPR. ENS - 48, boulevard Jourdan - 75014 Paris, France. Thierry.Magnac@ens.fr.
http://www.inra.fr/Internet/Departements/ESR/UR/lea/magnac.htm

‡LEA-INRA and CREST-INSEE. . Also affiliated with the CEPR. ENS - 48, boulevard Jourdan - 75014 Paris, France. Jean-Marc.Robin@ens.fr.
http://www.inra.fr/Internet/Departements/ESR/UR/lea/robin.htm.

§Financial support from the Commissariat Général du Plan and NATO (Combes' advanced fellowship grant) are gratefully acknowledged. We thank Eve Caroli and Eric Maurin for comments on this work.

1 Introduction

Huge amounts of efforts and money are spent in Europe to reduce regional inequalities. In the U.S., labor mobility is often seen as high enough to make inequalities a secondary problem although the recent increase in regional disparities also leads to the revival of this issue. One of the fundamental questions that emerges in this ongoing debate is to determine whether regional inequalities rapidly and spontaneously vanish or are persistent and have to be tackled through redistribution policies. A related question concerns the type of competition policies that local authorities might adopt to make firms grow or to induce new firms to be created. Our paper contributes to this debate by providing empirical evidence on the evolution and determinants of local employment in a large European country, France.

The literature on regional income convergence (the "β-convergence", see Barro and Sala-I-Martin [1995]) is prolific. However, it is not the most relevant theoretical background for the present study. Whereas the geographical level considered in these approaches is most often the U.S. states or the NUTS 2 regions in Europe, we consider much smaller geographical units (341 units for France, equivalent to NUTS 4). This choice has two major consequences. First, our study more likely belongs to the urban and economic geography literature than to the empirical literature about growth. These are more relevant to explain intra-regional disparities such as the ones between cities or between rural and urban areas for instance. Second, while β-convergence focuses on the determinants of output and income, we concentrate on the determinants of local inequalities in employment and employment growth in which people and decision makers are indeed more often interested in Europe due to the large persistence of unemployment.

This study finds its roots in the seminal papers by Glaeser, Kallal, Scheinkman and Schleifer [1992] and Henderson, Kuncoro and Turner [1995]. These authors try to link local long-run industrial employment growth to local industrial specialization (the so-called "MAR externalities") and diversity ("Jacobs externalities"). Glaeser *et al.* [1992] also consider the impact of local competition ("Porter" externalities). As surveyed by Combes and Overman [2004], this has been replicated on various countries[1], but contradictory results are sometimes obtained. This can be explained by differences across countries or time periods in the strength of agglomeration forces, but also by some methodological issues.

Under this light, another resourceful study is Henderson [1997]. He shows that the most important of these drawbacks can be tackled by using panel data methods to model the short-run dynamics of local employment and the impact of the local economic structure. First, this allows to distinguish possible local area and industry fixed effects (capturing for instance the role of physical geography, climate, endowments, culture, etc) from local and time dependent externalities linked to the local economic structure. Furthermore, endogeneity, which may be

[1]Among which France by Combes [2000] using the same data as in the present study.

critical, can be properly controlled for. Second, this allows to state whether externalities have a long lasting impact ("history matters") and correspond to dynamic externalities or influence local growth only in the short run and are mainly static. This old debate in urban economics is critical in terms of policies: Altering the local economic structure should be much more rapidely efficient in the second case.

Our approach also relies on dynamic panel data analysis in order to distinguish between fixed effects and economic structure and between static and dynamic externalities simultaneously controlling for endogeneity. As it is based on French data, it provides interesting comparisons between the European and the U.S. regional dynamics. Yet, it sensibly differs from previous studies. First, we employ model selection techniques and parsimonious representations of the series, which lead to select a model with a lag structure of a low order (ARMA(1,1)). Second, the effect of local specialization is most often identified in the literature thanks to a non-linear effect of sectoral employment that enters the specification both in logarithm and level. To put it another way, if the model were specified only in logarithm, the specialization effect could not be identified from the initial sectoral employment effect (Combes, [2000]). This makes interpretations problematic when both effects act in opposite directions as in Henderson *et al.* [1995], the reason why we prefer to not use this specification. Besides, we consider in the explanatory variables simultaneously two measures of diversity that add to total employment to capture urbanization economies and two indices of local competition to study Porter externalities.

Last and maybe most importantly, a critical difference with former studies stems from decomposing local industrial employment into the product of average plant size and number of plants. Instead of working with the dynamics of local industrial employment only, we simultaneously study the dynamics of these both variables as embodied in a Panel Vector Autoregression (PVAR) setting. Local growth is therefore decomposed into "internal" local growth defined as the growth of the size of existing plants in the area, and "external" local growth defined as the creation of new plants in the area. For each component, we allow for different dynamics and determinants, which provides new insights into the local growth factors. This decomposition is dictated by a simple theoretical argument developed in the next section of this paper. Under imperfect competition, local externalities simultaneously increase the size of existing plants and drives new firms into the market.

We use data on employment in 36 industries covering manufacturing, trade, and services, available for the 341 French continental "Employment Areas" (EAs). The period of observation covers 10 years from 1984 to 1993. The low order of the model selected implies that static externalities are prevalent compared to dynamic ones, contrary to the U.S. (see Henderson, 1997). This would make regional policies more rapidly efficient but also less lasting. The autocorrelation coefficient is larger for plant size than for the number of plants. Moreover, whereas fixed effects explained most of the spatial variation of plant size, plant creation is mainly determined by the current local economic structure. For both reasons, policies targeted on the creation of new plants should prove to be more efficient. We show that the existence of agglomeration economies is linked to the size of the local economy: Larger areas experience larger growth in both the number of plants and the plant size. This can be attributed to gains due to demand and cost linkages (market-based agglomeration forces) or to strong/high quality technological spillovers (pure externalities) that are not industry specific. We also refine the verdict regarding the impact of the industrial composition of the local economy on local growth. Whereas the number of locally active sectors does not need to be large, sectors of comparable size favor both internal and external growth. An interpretation would be that technological spillovers can work across industries but do not extend to all sectors. Similarly, intermediate inputs at the origin of demand and cost linkages would be not necessary numerous in each industry but equally important. Hence, in both cases, the optimal structure would be small groups of even size sectors. Last, as regards the impact of local competition, it is shown to be non linear. For a given size of the local economy, plants appear to be larger in areas where they are more numerous, definitively not in a situation of local monopoly, but of uneven size. Large leaders, either relying on economies of scale or having research and development units of efficient size, would impulse growth to smaller and numerous plants surrounding them. Those smaller plants benefit for instance from technological spillovers or from large markets and from improved matching with their partners. On the other hand, the number of plants grows faster in places where plants are less numerous and of even size. All of these clearly shed new light on the local economic structure that is the most favorable to local growth in a set-up we believe to be robust and easy to replicate.

Section 2 is devoted to a clarification of the theoretical background. Next, section 3 presents the data we use in our application and analyzes some descriptive statistics concerning the sample structure as well as a simple covariance analysis. Estimation results of static and dynamic PVAR models of plant size and number of plants are presented in section 4. They are interpreted and discussed in section 4.3, while section 5 concludes and proposes new lines of research.

2 Economic Background

The lack of a precisely identified background model is one of the drawbacks of the studies linking local industrial growth to local economic structure. It is a hard task indeed to provide a rigorous framework to these estimations that, in our opinion, have to be viewed as proposing stylized facts and not as validating a given theory. For instance, local growth of a region depends on the characteristics of other regions in an intricate way as soon as inter-regional trade is considered. Therefore, for the sake for simplicity, an important implicit assumption in this literature is that each region is a closed economy, the local growth of which is linked to the economic structure of this very region only. There is a critical need of clarifying all the simplifying assumptions made is this literature. We develop in this section a simple framework that helps identifying them. Next, we show how it can be extended to integrate new features such as imperfect competition and firm creation.

2.1 A Competitive Set-up

Let us consider a setting in which each region z is a closed economy. Only one good is produced, under constant returns to scale, using labor, L_z, and capital, K_z. Production, Y_z, is assumed to be given by:

$$Y_z = A_z (L_z)^{\alpha} (K_z)^{1-\alpha}, \tag{1}$$

where α is a constant between 0 and 1, and A_z is total factor productivity. If we assume perfect competition both on the good and factor markets, the equilibrium price is obtained as:

$$p_z = \frac{(w_z)^{\alpha} r^{1-\alpha}}{\alpha^{\alpha} (1-\alpha)^{(1-\alpha)} A_z}, \tag{2}$$

where w_z is the wage and r the return to capital. Labor markets are supposed to be local and labor mobility imperfect, the reason why the wage is region specific. By contrast, capital markets are assumed to be perfectly integrated.

Let us focus our attention on the impact of a productivity shock, $\frac{dA_z}{A_z}$, on employment growth, $\frac{dL_z}{L_z}$. Next, it is assumed that productivity shocks depend on local characteristics such as specialization, diversity, competition, or total region size in order to determine the impact of these local characteristics on growth. Let σ denote the demand elasticity of the good, and ε the supply elasticity of labor. ε is assumed to be a finite positive constant and, by contrast, the capital supply elasticity is supposed to be infinite, assumptions consistent with the previous ones regarding mobility. By definition, we have:

$$\frac{dY_z}{Y_z} = -\sigma \frac{dp_z}{p_z} \text{ and } \frac{dL_z}{L_z} = \varepsilon \frac{dw_z}{w_z}. \tag{3}$$

Differentiating equations (1) and (2) and equalizing labor productivity and real wages, simple algebra developed in Appendix A.1 leads to:

$$\frac{dL_z}{L_z} = \varepsilon (\sigma - 1) \mu \frac{dA_z}{A_z}, \tag{4}$$

where:

$$\mu = \frac{1}{1 - \alpha + \varepsilon + \alpha\sigma}$$

is a positive parameter.

Many interesting points are already worth noting as regards employment growth following a productivity shock. First, a positive productivity shock has a positive effect on employment if and only if the demand elasticity is greater than 1. It is intuitive that if demand does not expand enough after the price decrease, employment shrinks since productivity enhancements save inputs. The larger the demand elasticity, the larger the production growth, and thus the more employment expands. Next, the larger the labor supply elasticity, the more employment grows. At one extreme if $\varepsilon = 0$, employment does not expand since no additional supply of labour is locally available even for higher wage. The productivity gain translates into an increase of output thanks to an increase in capital only. Otherwise, the larger the labor supply elasticity, the lower the wage increase, the larger the price decrease, and thus the more output and next labor expand. If the labor supply elasticity is infinite (local non voluntary unemployment situation for instance), the wage does not change and the effect on output and employment is maximized, equal to $\sigma - 1$.

Therefore, this simple framework leads to clarifying some important aspects. The impact of local externalities on employment growth is positive only if the demand elasticity is high enough and the larger the labor supply elasticity, the larger it is. Previous literature assumes that a positive productivity shock always induces employment expansion.

Glaeser *et al.* [1992] introduce an index of local competition among explanatory variables. This can be interpreted as a direct effect of competition on the strength of externalities, and therefore on employment growth through the productivity shock. In the perfect competition setting they assume, the number of competitors is not defined however. Moreover, competition may also simultaneously play a critical role on employment through its impact on firm mark-ups and next on price and production levels if competition is imperfect. Hence, it seems relevant to slightly extend the previous set-up to better understand the role of competition.

2.2 Imperfect Competition

It leads to simple and intuitive results to assume that the good is homogenous and that imperfect competition is modelled as a static Cournot equilibrium. Consider first the short-run situation in which the number of (single-plant) firms located in the region, N_z, is exogenous. Each firm maximizes its profit with respect to the quantity it produces, taking into account the non-zero demand elasticity and assuming that other firms hold constant their own output. As developed in Appendix A.2, the equilibrium price is:

$$p = \frac{\sigma N_z}{\sigma N_z - 1} \cdot \frac{(w_z)^\alpha r^{1-\alpha}}{\alpha^\alpha (1-\alpha)^{(1-\alpha)} A_z}, \tag{5}$$

that is to say, a mark-up that depends on the number of firms over the marginal cost. The production function and the equalization of the real wage to productivity can again be differentiated along with the additional equation (5) to obtain the employment dynamics:

$$\frac{dL_z}{L_z} = \varepsilon (\sigma - 1) \mu \left(\frac{dA_z}{A_z} + \frac{1}{\sigma N_z - 1} \frac{dN_z}{N_z} \right). \tag{6}$$

If the number of firms remains fixed, $\frac{dN_z}{N_z} = 0$, the effect is exactly the same as under perfect competition, and local competition, here embodied in the number of local firms, affects the local employment growth only through its possible impact on the productivity shock, $\frac{dA_z}{A_z}$.

Nevertheless, since in the short-run profits are affected by productivity shocks, one may expect some firms to be simultaneously created and enter the market. In this case an extra effect is at work. An increase in the firm number, $\frac{dN_z}{N_z} > 0$, has a negative impact on price. If the demand elasticity is greater than 1, this induces a positive impact on employment, which reflects in the additional term in equation (6) compared to equation (4).

At this point, it is worth decomposing local employment growth into the sum of two terms: First, an effect we call "internal growth" which is the growth in size of existing firms, $\frac{dl_z}{l_z}$, where l_z is the average firm size in area z. The second effect concerns "external growth", which is the expansion of the number of firms, $\frac{dN_z}{N_z}$. In other words, we simply write:

$$\frac{dL_z}{L_z} = \frac{dl_z}{l_z} + \frac{dN_z}{N_z}, \tag{7}$$

Using equation (6), the impact of local externalities on firm size, $\frac{dl_z}{l_z}$, is given by:

$$\frac{dl_z}{l_z} = \varepsilon \left(\sigma - 1\right) \mu \left(\frac{dA_z}{A_z} + \frac{1}{\sigma N_z - 1} \frac{dN_z}{N_z} \right) - \frac{dN_z}{N_z}. \tag{8}$$

If the number of firms does not change, firm size grows at the rate of total employment in the perfect competition setting, that is to say only through the channel of productivity on prices. If the number of firms increases, firm size increases through both this effect and the indirect impact of competition on prices as they appear in the first two terms of equation (6). Besides firm size simultaneously decreases through a direct competition effect, the last term in equation (8). Total output is split between a larger number of firms, each of them becoming smaller if the impact on total employment is not strong enough. Therefore, the total effect of the productivity increase on firm size can be negative if many firms simultaneously enter into the local market.

Next, we can separately study the impact of local externalities on firm creation, $\frac{dN_z}{N_z}$. In the previous discussion, we implicitly assume that the firm creation process is exogenous. Local externalities may have an endogenous impact on the number of plants, however. This is the case for instance when the number of firms adjusts to a long-run equilibrium such as profits are equal to zero.[2] By plugging the equilibrium price given by equation (5) in the profit definition and by next differentiating this equation, we obtain the adjustments in firm size and firm number following a productivity shock in the long-run situation (see appendix A.3). They are given by:

$$\frac{dl_z}{l_z} = \left(\varepsilon - 1\right) \left(\sigma - 1\right) \lambda \frac{dA_z}{A_z}, \tag{9}$$

[2] The assumption of fixed costs paid at every period makes this number well defined (see Appendix A.3).

where:

$$\lambda = \frac{(\sigma N_z - 1)}{2\alpha(\sigma - 1)(\sigma N_z - 1) + (1+\varepsilon)(\sigma(N_z - 1) + \sigma N_z - 1)}$$

is positive. The variation of the firm number is:

$$\frac{dN_z}{N_z} = (1+\varepsilon)(\sigma - 1)\lambda \frac{dA_z}{A_z}. \tag{10}$$

The total impact of the productivity shock on total employment can be recovered as:

$$\frac{dL_z}{L_z} = 2\varepsilon(\sigma - 1)\lambda \frac{dA_z}{A_z}. \tag{11}$$

Therefore, all variables which have an impact on productivity also have an impact both on internal and external growth. Moreover the effects have different magnitude on both variables, and even possibly different signs.

In the long-run, a positive productivity shock has a positive effect on firm creation if the demand elasticity is sufficiently high (equation 10). σ greater than 1 insures that the price decrease effect on demand is large enough to compensate the better efficiency of labor and therefore to induce both an increase in total employment and firm number.

As regards firm size, the same result holds if the labor supply elasticity is large enough: The positive indirect effect of the increase in competition (due to the price decrease) dominates the direct negative effect (the last term in equation 8). If the labor supply elasticity is low, firm size in the long run equilibrium decreases following a positive productivity shock.

It would be possible to consider less extreme cases, such as for instance a situation in which profits are different from zero and firm creation is proportional to them. Weaker competition settings as for instance price competition with differentiated goods would also reduce the direct competition effect at the cost of more tedious algebra. On the other hand, it would introduce more flexibility into the model. Another interesting direction to explore would be given by a setting where shocks also affect recurring fixed costs of entry paid by firms at each period. The structure of local production could indeed affect these fixed costs and therefore entry into the local market. The proportionality between coefficients in equations (9) and (10) would then likely be broken.

Even if these possible extensions could be worth studying, we do not need them here. Indeed, the productivity shock may itself depend on the competition degree (Porter externalities). This is sufficient to lead us to study the possibly contrasted impact of the local economic structure on average firm size and the number of firms, embodied in a VAR model. It simultaneously models both variables in the system of equations (9) and (10) or equivalently of equations (8) and (10). As a matter of fact, it turns out that the latter system of equations is more convenient for estimation purposes because it forms a recursive system. It is also interesting to use equation (8) instead of equation (9) since it does not depend on the assumption that the number of firms instantaneously adjusts to the long run equilibrium or not. Notice that the simultaneous estimation of firm size and number of firms dynamics allows us to identify more effects than usually done in the literature. We are able to distinguish the impact of the local economic structure on internal and external growth separately and not only the aggregate impact on total growth.

Finally the candidate models which are able to sustain the kind of estimations performed in this literature necessarily assume that each region is an economy facing given demand and supply functions that are independent of what happens in the neighbouring regions. Indeed, as soon as trade between regions is properly modelled for instance, the equilibrium price depends not only on local productivity but also on productivity of all trading partners. Derived specifications, as for instance Combes and Lafourcade [2001] for Cournot competition, are much more intricate and beyond the scope of this literature. Again, the purpose is to provide stylized facts on local growth dynamics and not to estimate a precise given model of trade and technology diffusion for instance.

3 A Brief Look at the Data

We now describe the data we use and the variables that enter the VAR models. Data come from a survey on plant employment ("*Enquête Structure des Emplois*") collected by the French National Institute of Statistics and Economic Studies (INSEE). The survey reports the employment level of each plant between 1984 and 1993, the area where the plant is located, and its industry. We prefer to use data on plants rather than firms since the actual location of production is therefore better localized. Plants within a firm are assumed to be independent in terms of technological diffusion and production choices, and no internal labor markets are supposed to exist. By doing this we neglect technological diffusion and job mobility within firms because we feel that richer data on firms should be used to analyse externalities at the firm level. Externalities we consider are therefore limited to the local level. One of the drawback of the survey is that it includes all French-metropolitan plants that employ more than 20 employees only. Comparable truncations are also common on U.S. city data. We leave for future research the assessment of whether our results are affected by this selection of plants.

Data from individual plants could be used as they are. Nevertheless, growth determinants are area and industry specific and individual characteristics play no role in our model. Furthermore, whereas plant size is a variable that is easy to model using individual plant data, modeling the increase in the number of plants increases the level of difficulty. For these reasons we prefer to aggregate employment across small local areas and across industries at a rather detailed level of aggregation. We exclude agriculture and non-market services and group activities into 36 different manufacturing, trade, and service industries. The area code distinguishes 341 geographic units or employment areas ("*zones d'emploi*"). This INSEE spatial nomenclature exhaustively covers the French territory, thus corresponding to both rural and urban areas. The average area spreads over 1,570 km^2, which is fairly small (equivalent to splitting the continental U.S. territory in more than 4,700 units). The employment area definition is based on the observation of workers' daily migrations. Given the number of areas, it attempts to maximize the number of within-area migration between home and work. Hopefully, this maximizes the coincidence between the geography of where people live and the geography of where people work, that is to say between local labor markets and local good markets. Importantly, this is consistent with the assumption that local growth only depends on local characteristics.

3.1 Dependent Variables

The local employment structure of area z and sector s at time t is characterized by the pair of variables $(L_{zst}/N_{zst}, N_{zst})$, where L_{zst} is employment in area z and sector s at time t and where N_{zst} is the number of plants located in area z and operating in sector s at time t. We adopt a logarithmic specification. It has the double advantage of making the distribution of these variables closer to a normal distribution and of making easier the interpretation of first-differences as growth rates.

We begin the description of the data with a rough view of the sample structure. Table 1 reports the number of observations (i.e. aggregates of plants with more than 20 employees in area and sector (z, s) at time t) that have been "active" (i.e. $L_{zst} > 0$) since year $t-1$, $t-2$ and so on till $t = 1$, the first observation period (1984).[3] Local areas and sectors where plants started operating after 1984 are scarce except in 1985 when some areas of Provence were surveyed inadequately. Entries just below the diagonal reflect that problem. Yet, most observations (z, s) are such that employment L_{zst} is positive at the first date ($t = 1984$) and the panel is therefore approximately balanced. We neglect entries and exits (attrition) of areas and sectors into the panel in the empirical analysis. First, because the bulk of such movements comes from the survey problem in Provence that we may presume to be exogenous to the size and number of plants (it was a coding error according to INSEE). Second, even if the endogeneity of entries and exits into the sample is obvious because they are commanded by the size of a firm becoming larger or smaller than 20 employees, the number of such cases is small (see Table 1). Furthermore, when Tobit corrections are applied to simpler contexts, the effect is far from dramatic (Combes [2000]). Finally, we can always interpret our results conditional on areas and sectors being in the sample since in the absence of a structural estimation our results are mainly descriptive.

[3] The table reports the number of observations (z, s, t) for any possible pair (t, τ_{zst}) where $\tau_{zst} = t - \min\{t | L_{zst} > 0\} + 1 \in \{1, ..., t\}$.

Table 1: Numbers of active area/industry pairs in the sample at each date

year	total	1	2	3	4	5	6	7	8	9	10
1984	7786	7786									
1985	8322	719	7603								
1986	8341	148	719	7474							
1987	8384	123	160	732	7369						
1988	8349	86	112	158	732	7261					
1989	8292	81	76	116	161	725	7133				
1990	8328	88	96	74	116	176	733	7045			
1991	8287	68	75	88	75	110	179	721	6971		
1992	8248	61	64	78	82	79	105	170	700	6909	
1993	8251	72	55	70	80	84	88	110	182	684	6826

Note : Entries in the table are read as: In 1984, the sample is composed by 7786 pairs (z, s) of local areas and sectors; In 1985, 719 pairs (z, s) entered the sample and were not there in 1985 while 7603 pairs (z, s) are active and were in the sample in 1984, for a total of 8322, etc.

3.2 Explanatory Variables

Besides MAR externalities modelled by the level of persistence in the series, two groups of determinants of local growth make up the list of explanatory variables. The first group contains variables measuring urbanization externalities that describe total local employment and sectoral diversity of plants in the local market (Jacobs externalities). The variables in the second group measure Porter effects that characterize the magnitude of competition between plants in the same sector. We specifically use the following indices in our study:

1. The logarithm of total employment in area z at date t :

$$lL_{zt} = \log \left[\sum_{s=1}^{S} L_{zst} \right].$$

   This first variable, frequently used in the literature, captures global urbanization externalities, those only related to market size and not to the activity composition in terms of industries.

2. The logarithm of the number of sectors, S_{zt}, in which at least one plant (employing more than 20 workers) is operating in area z at date t :

$$ls_{zt} = \log(S_{zt}).$$

3. The opposite of the Herfindahl index of local concentration between sectors:

$$ld_{zt} = -\log \left[\sum_{s=1}^{S} \left(\frac{L_{zst}}{L_{zt}} \right)^2 \right],$$

This variable is equal to zero if local employment is concentrated into a single sector and it is equal to the logarithm of the number of sectors if the distribution of local employment is uniform across sectors.

These last two indicators measure the sectoral diversity of an employment area. They correspond to what is usually called Jacobs externalities, which constitutes the second kind of urbanization externalities.

Urbanization externalities variables are specific to area z and date t only, but not to industry. We now move on to the local competition variables that vary across area z, sector s, and date t.

4. The dispersion of local employment between plants within a sector as measured by the opposite of the logarithm of the Herfindahl index of within-sector and within-area concentration:

$$lcom_{zst} = -\log\left[\sum_{i\in(z,s,t)}\left(\frac{\ell_{it}}{L_{zst}}\right)^2\right],$$

where ℓ_{it} is employment in the i^{th} plant at period t and where we denote $\{i \in (z,s,t)\}$ the set of all plants i operating in area z and sector s at period t.

If employment is concentrated in a single plant, this variable is equal to zero. It is equal to the logarithm of the number of plants if the distribution of employment is uniform among plants. Given the number of plants, we interpret this variable as the intensity of local competition within sectors (Encaoua and Jacquemin, 1980).

5. An indicator of total absence of competition within an area and sector:

$$\begin{aligned} c_{zst} &= 1 \text{ if } N_{zst} = 1, \\ &= 0 \text{ if not.} \end{aligned}$$

Our specification regarding competition is therefore partly non-linear since this monopoly variable depends on the second dependent variable, the logarithm of the number of plants. As a consequence, absence of competition is included among the explanatory variables in the plant size equation but it is excluded from explanatory variables in the equation describing the number of plants.

This choice of variables is justified by the survey of the literature presented in the introduction and the model detailed in section 2. It is worth mentioning that the usual index of specialization, which is the ratio of employment in area z and industry s over total employment in this area (L_{zst}/L_{zt}), is not retained here. In logarithms, the effect of this variable would not be parametrically identified because of the collinearity between the dependent variables $\log(L_{zst}/N_{zst})$

and $\log(N_{zst})$ and the market size indicator $\log L_{zt}$ (see Combes [2000]). MAR externalities are captured by the lagged dependent variables introduced in the explanatory variables. Furthermore, we extend the standard approach by considering simultaneously different measures of local diversity and competition.

Table 2 provides descriptive statistics for all variables. In particular note that in more than 20% of cases, labor employed by all plants with more than 20 employees in an area and sector is employed by a single plant, (c_{zst}). Besides, the variability in the second dependent variable is larger than in plant size.

Table 2: Descriptive statistics

	Av.	Std.	Min	Max
y^1_{zst}	4.18	0.76	2.99	10.12
y^2_{zst}	1.49	1.16	0	7.54
lL_{zt}	9.51	1.08	6.51	13.59
ls_{zt}	3.22	0.24	1.79	3.58
ld_{zt}	2.37	0.42	0.34	3.12
$lcom_{zst}$	1.16	0.95	0	6.33
c_{zst}	0.21	0.41	0	1

Notes: **(a)** There are 82853 cells (z, s, t) in which employment L_{zst} is strictly positive. **(b)** *Variables definition:* y^1_{zst}: log(plant average employment); y^2_{zst}: log(number of plants); lL_{zt}: log(total employment in the area); ls_{zt}: log(number of sectors with at least a plant with more than 20 employees in the area); ld_{zt}: log(index of local diversity between sectors); $lcom_{zst}$: log(index of local competition within sectors); c_{zst}: index of monopoly situation at the local and sectoral level.

Time and industry dummies explain about one fourth of the total variance of variables $\log(L_{zst}/N_{zst})$ and $\log N_{zst}$ (25% and 23% respectively). As we are not interested here in analyzing the effect on local employment dynamics of macro- and industry-specific shocks but aim at characterizing spatial effects, we substract from all variables their mean within period and industry cells. The econometric analysis therefore exclusively applies to the vector $\mathbf{y}_{zst}$ of dependent variables, defined as:

$$\mathbf{y}_{zst} \equiv \begin{pmatrix} y^1_{zst} \\ y^2_{zst} \end{pmatrix} = \begin{pmatrix} \log\left(\dfrac{L_{zst}}{N_{zst}}\right) \\ \log N_{zst} \end{pmatrix} - \frac{1}{\#\{z \in (s,t)\}} \sum_{z\in(s,t)} \begin{pmatrix} \log\left(\dfrac{L_{zst}}{N_{zst}}\right) \\ \log N_{zst} \end{pmatrix},$$

and the vector $\mathbf{x}_{zst}$ of conditioning variables, defined as:

$$\mathbf{x}_{zst} = \begin{pmatrix} lL_{zt} \\ ls_{zt} \\ ld_{zt} \\ lcom_{zst} \\ c_{zst} \end{pmatrix} - \frac{1}{\#\{z \in (s,t)\}} \sum_{z\in(s,t)} \begin{pmatrix} lL_{zt} \\ ls_{zt} \\ ld_{zt} \\ lcom_{zst} \\ c_{zst} \end{pmatrix},$$

denoting $\#\{\cdot\}$ the number of elements of the set $\{\cdot\}$.

3.3 Contemporaneous Correlations between Variables

Table 3 reports correlations between dependent variables, $\mathbf{y}_{zst}$, and determinants, $\mathbf{x}_{zst}$. First, the larger plant size (y^1_{zst}), the larger the number of plants (y^2_{zst}). Second, the larger plant size (y^1_{zst}), or the larger the number of plants (y^2_{zst}), the larger the local size (lL_{zt}), the number of active sectors (ls_{zt}), the diversity between sectors (ld_{zst}), and the local competition within sectors ($lcom_{zst}$), and the less likely a monopoly situation (c_{zst}). These correlations seem to reflect mainly the contrast between small and large markets.

To abstract from size effects, we report correlations between growth rates in table 4. As is often the case with panel data, correlations between growth rates are generally weaker than those in level. Some correlations are quite significant, however. For instance, correlations between, on the one hand, the dependent variables (plant size and the number of plants) and local diversity and total employment on the other hand remain positive, while the correlation between plant size and local competition within sectors becomes negative.

Table 3: Correlations between variables in levels

(Obs 82853)	y^1_{zst}	y^2_{zst}	lL_{zt}	ls_{zt}	ld_{zt}	$lcom_{zst}$	c_{zst}
y^1_{zst}	1						
y^2_{zst}	0.278	1					
lL_{zt}	0.205	0.717	1				
ls_{zt}	0.167	0.595	0.840	1			
ld_{zt}	0.100	0.351	0.358	0.594	1		
$lcom_{zst}$	0.121	0.947	0.668	0.563	0.335	1	
c_{zst}	-0.218	-0.608	-0.352	-0.330	-0.222	-0.569	1

Note: *Variables definition:* See table 2, note (b).

Table 4: Correlations between first-differenced variables

(Obs 72780)	y^1_{zst}	y^2_{zst}	lL_{zt}	ls_{zt}	ld_{zt}	$lcom_{zst}$	c_{zst}
y^1_{zst}	1						
y^2_{zst}	-0.095	1					
lL_{zt}	0.081	0.213	1				
ls_{zt}	0.004	0.088	0.387	1			
ld_{zt}	0.026	0.083	-0.069	0.250	1		
$lcom_{zst}$	-0.264	0.828	0.155	0.077	0.069	1	
c_{zst}	0.032	-0.677	-0.082	-0.039	-0.047	-0.563	1

Note: *Variables definition:* See table 2, note (b).

4 Econometric Analysis

The univariate statistical analysis reported in Appendix B suggests the use of a vector autoregressive specification such as:

$$\mathbf{y}_{zst} = A_0(L)\mathbf{y}_{zs,t-1} + B_0(L)\mathbf{x}_{zst} + \widetilde{\varepsilon}_{zst} \tag{12}$$

where $A_0(L)$ and $B_0(L)$ are matrix polynomials in the lag operator L and where $\widetilde{\varepsilon}_{zst}$ is a vector of random shocks. Dependent variables are correlated and the variance-covariance matrix of $\widetilde{\varepsilon}_{zst}$ is not supposed to be a diagonal matrix though it is supposed to be constant over time.

It is always possible to rewrite system (12) using one of its recursive forms (see for instance Gouriéroux and Monfort, 1999):

$$y^1_{zst} = A_{11}(L)y^1_{zs,t-1} + A_{12}(L)y^2_{zst} + B_1(L)\mathbf{x}_{zst} + \eta^1_{zst} \tag{13}$$

$$y^2_{zst} = A_{21}(L)y^1_{zs,t-1} + A_{22}(L)y^2_{zs,t-1} + B_2(L)\mathbf{x}_{zst} + \eta^2_{zst} \tag{14}$$

where random shocks η^1_{zst} and η^2_{zst} are now uncorrelated and where $A_{ij}(L)$ and $B_i(L)$ are scalar polynomials in the lag operator. We chose this recursive form to emphasize that plant size should be varying at higher frequency than the series recording the number of plants. It is also justified by the theoretical argument that employment decisions are taken conditional on the entry decisions of plants decided beforehand. Equation (13) corresponds to the structural equation (8) describing changes in plant size as a function of productivity shocks and the number of plants. Equation (14) corresponds to the other structural equation (10) when the condition that profits are equal to zero is used to model plant creation.

As shocks are now uncorrelated, we can separately estimate the two equations using the same methodology. To improve reading we insist in the next two sections on model selection only. These "econometric" sections are then followed by section 4.3 dealing with economic interpretations.

4.1 Average Plant Size

Two modeling frameworks are worth exploring empirically. One model is "static" in the sense that polynomials $A_{ij}(L)$ and $B_i(L)$ are constant and A_{ii} are supposed to be equal to zero. Second, in the so called dynamic model, these polynomials are supposed to be of higher order and, as a result of our empirical analysis, it is likely that they are of order 1.

4.1.1 Static Models

Given the previous discussion and the univariate analysis presented in appendix B, we use the following specification:

$$\begin{aligned} y^1_{zst} &= \alpha y^2_{zst} + \mathbf{x}'_{zst}b + u_{zs} + \varepsilon_{zst} \\ &\equiv \widetilde{\mathbf{x}}'_{zst}b + u_{zs} + \varepsilon_{zst} \end{aligned} \tag{15}$$

where u_{zs} stands for an area and sector effect. In particular, as we do not have dynamic information on local infrastructure or on the local availability and type of local public goods, we assume that their effects are translated into these area and industry effects or, if they evolve over time, in the shock ε_{zst}. Area and industry effects could also stand for the more traditional local endowment advantages, the availability of specific inputs for instance or even for the local physical geography.

In the case where covariates are exogenous ($\widetilde{\mathbf{x}}_{zst}$ and ε_{zst} are uncorrelated) and are uncorrelated with area and sector effects, u_{zs}, OLS yields consistent estimates. Results are reported in the column 1 of table 5. The coefficient of determination (R^2) is quite large (27.4%). As is well known however, OLS estimates are biased if explanatory variables $\widetilde{\mathbf{x}}_{zst}$ are correlated with area and sector effects u_{zs}. It is why we report in the column 2 within-estimates which are robust to the former specification error under the assumption that variables $\widetilde{\mathbf{x}}_{zst}$ are strongly exogenous (x_{zst} and $\varepsilon_{zst'}$ are uncorrelated at any dates t and t'). Taking first differences and estimating by OLS the differenced series is another method to eliminate area and sector effects which results are reported in the column 3. These two procedures give very similar results.

Nevertheless, serial dependence is shown to be important in the analysis of univariate series: In table B.1 in appendix B, all autocorrelations of residuals are significantly different from zero, whatever the variable. Moreover, the persistence in the series cannot be explained by area and sector effects only since area and sector effects are controlled. Therefore, explanatory variables are not strictly exogenous: They are weakly exogenous only ($\widetilde{\mathbf{x}}_{zst}$ and $\varepsilon_{zst'}$ are uncorrelated at all dates t less or equal to t') or quasi-weakly exogenous ($\widetilde{\mathbf{x}}_{zst}$ and $\varepsilon_{zst'}$ are uncorrelated at all dates t strictly less than t')). It is the case when past random shocks on the dependent variable only affect the future and/or present of predetermined explanatory variables.

When explanatory variables are weakly exogenous, first-differentiating the model eliminates area and sector effects, u_{zs}, while the serial dependence structure remains simple (in contrast with a within-type transformation). This corresponds to the model written as:

$$\Delta y^1_{zst} = \Delta \widetilde{\mathbf{x}}'_{zst} b + \Delta \varepsilon_{zst}.$$

When variables are quasi-weakly exogenous, the lagged explanatory variables at orders 2, 3, etc can be used as instruments in this equation (Hausman and Taylor, 1981). 2SLS estimation results are reported in column 5 of table 5. When variables are weakly exogenous, lagged variables at orders 1, 2, etc are valid instruments. 2SLS estimation results are reported in column 6.

Alternatively, one might consider the initial equation in levels:

$$y^1_{zst} = \widetilde{\mathbf{x}}'_{zst} b + u_{zs} + \varepsilon_{zst},$$

and use as instruments the first differences $\Delta\widetilde{\mathbf{x}}_{zst}$, $\Delta\widetilde{\mathbf{x}}_{zst-1}$,..., if $\widetilde{\mathbf{x}}_{zst}$ is weakly exogenous or $\Delta\widetilde{\mathbf{x}}_{zs,t-1}$, $\Delta\widetilde{\mathbf{x}}_{zst-2}$, etc (i.e. omitting from the list $\Delta\widetilde{\mathbf{x}}_{zst}$) if variables are quasi-weakly exogenous

Table 5: Average plant size - Static model

	(1)	(2)	(3)	(4)	(5)	(6)
Method	OLS	Within	OLS	2SLS	2SLS	2SLS
y^1_{zst}	L		Δ	L	Δ	Δ
Instruments	-	-	-	$(\Delta \mathbf{x}_{zs,t-j})_{j=1,2,3}$	$(\mathbf{x}_{zs,t-j})_{j=2,3,4}$	$(\mathbf{x}_{zs,t-j})_{j=1,2,3}$
y^2_{zst}	1.000	0.257	0.283	0.302	0.138	0.327
	(148.5)	(36.9)	(43.7)	(4.9)	(1.9)	(14.7)
lL_{zt}	-0.016	0.311	0.277	0.052	0.173	0.226
	(-3.8)	(31.8)	(30.4)	(0.4)	(1.3)	(6.5)
ls_{zt}	0.052	-0.203	-0.183	0.550	0.041	-0.277
	(2.7)	(-10.6)	(-10.5)	(3.0)	(0.2)	(-4.5)
ld_{zs}	0.003	0.148	0.137	-0.540	0.096	0.143
	(0.51)	(16.3)	(15.1)	(-2.5)	(1.2)	(4.8)
$lcom_{zst}$	-1.116	-0.544	-0.566	-0.360	-0.469	-0.634
	(-148.2)	(-85.0)	(-94.1)	(5.1)	(-9.4)	(-34.2)
c_{zst}	-0.109	-0.066	-0.077	0.071	-0.074	-0.085
	(-16.2)	(-11.9)	(-15.7)	(1.5)	(1.4)	(-5.4)
R^2	27.4%	11.8%	13.2%	-	-	-
Sargan	-	-	-	$<10^{-5}$	0.030	0.038
AC(1)	0.906	0.954	-0.217	0.939	-0.206	-0.218
	$(<10^{-4})$	$(<10^{-4})$	$(<10^{-4})$	$(<10^{-4})$	$(<10^{-4})$	$(<10^{-4})$
AC(2)	0.852	0.931	-0.025	0.904	-0.029	-0.033
	$(<10^{-4})$	$(<10^{-4})$	$(<10^{-4})$	$(<10^{-4})$	$(<10^{-4})$	$(<10^{-4})$
AC(3)	0.806	0.912	-0.018	0.873	-0.035	-0.025
	$(<10^{-4})$	$(<10^{-4})$	(10^{-4})	$(<10^{-4})$	$(<10^{-4})$	$(<10^{-4})$
Obs	82846	82846	72773	46099	46099	54657

Notes : **(a)** *Variables definition*: See table 2, note (b). **(b)** *L and* Δ: Refer to variables in levels, Y_{zst}, or in first differences, ΔY_{zst}. **(c)** *Significativity*: Standard errors are not corrected for spatial or serial dependence. Students statistics are reported in brackets. **(d)** *Sargan*: This is the Sargan or J-test of overidentifying restrictions. p-values only are reported. AC(k) is the estimated value of the autocorrelation of order k. p-values associated to the test that these autocorrelations are equal to zero are reported below the estimates between brackets.

only. This estimation method of the model in level when the instruments are in first differences is, however, only consistent if explanatory variables $\widetilde{\mathbf{x}}_{zst}$ are stationary (see Blundell and Bond, 1998). With this caution in mind, we report in table 5, column 4 the two stage least squares estimates of the model in level using first differenced variables as instruments.

Given the relative merits of these estimation methods and specification statistics reported in the bottom part of table 5, our preferred results are in column 6. Namely concerning the test of overidentifying restrictions, the Sargan p-value is as large in column 6 as in column 5 and much higher than in column 4. Generally speaking, OLS results strongly contrast with all other results. It is sufficient to control area and sector effects by the within or first differenced OLS estimation methods (columns 2 and 3) to bridge most of the gap with our preferred results (column 6). At a lesser degree, note also that instrumenting levels by first-differences (column 4) differ from results of instrumenting first-differences by the variables in levels (columns 5 and 6), which might point out that processes on the RHS of the equation of interest are not stationary, as underlined by Blundell and Bond [1998]. The hypothesis that contemporaneous correlation between the dependent and explanatory variables is absent can be evaluated by contrasting columns 5 and 6 leading to comparable levels of the Sargan statistics, and it is not rejected. Instruments might be too weak, however, as shown by the imprecision of estimates in column 5.

Finally, specification diagnostics tell us that autocorrelation is significant up to the order 3 at least. This is why we are now looking for dynamic specifications that would agree with these results.

4.1.2 Dynamic Models

The most straightforward way to derive a dynamic model from a static equation like (15) is to assume that random shocks ε_{zst} follow an autoregressive process of order 1:

$$\varepsilon_{zst} = \rho\varepsilon_{zs,t-1} + (1-\rho)\eta_{zst},$$

where η_{zst} is stationary and is possibly autocorrelated. When $\rho < 1$, the process ε_{zst} is stationary. From equation (15) evaluated at periods t and $t-1$, we can derive that:

$$y^1_{zst} - \rho y^1_{zs,t-1} = \widetilde{\mathbf{x}}'_{zst}b + u_{zs} + \varepsilon_{zst} - \rho(\widetilde{\mathbf{x}}'_{zs,t-1}b + u_{zs} + \varepsilon_{zs,t-1}),$$

or equivalently:

$$y^1_{zst} = \rho y^1_{zs,t-1} + \widetilde{\mathbf{x}}'_{zst}b - \rho\widetilde{\mathbf{x}}'_{zs,t-1}b + (1-\rho)\left(u_{zs} + \eta_{zst}\right). \tag{16}$$

This expression is a particular case of the following linear model:

$$\begin{aligned} y^1_{zst} &= \rho y^1_{zs,t-1} + \widetilde{\mathbf{x}}'_{zst}b - \widetilde{\mathbf{x}}'_{zs,t-1}b_1 + (1-\rho)\left(u_{zs} + \eta_{zst}\right), & (17)\\ &= \rho y^1_{zs,t-1} + \widetilde{\mathbf{x}}'_{zst}(b - b_1) - \Delta\widetilde{\mathbf{x}}'_{zs,t-1}b_1 + (1-\rho)\left(u_{zs} + \eta_{zst}\right) & (18)\end{aligned}$$

when $b_1 = \rho b$.

Therefore, equation (16) is a constrained dynamic equation that agrees with a static relationship between the dependent and explanatory variables. This static relationship describes the long run equilibrium between these variables around which short run realizations evolve. Indeed, by iterating until $t = 1$, equation (16) can be written as:

$$y^1_{zst} = \widetilde{\mathbf{x}}'_{zst}b + u_{zs} + \rho^{t-1}(y_{zs1} - \widetilde{\mathbf{x}}'_{zs1}b - u_{zs}) + (1-\rho)(\eta_{zst} + \rho\eta_{zs,t-1} + ... + \rho^{t-2}\eta_{zs,2}),$$

in which the term $\rho^{t-1}(y_{zs1} - \widetilde{\mathbf{x}}'_{zs1}b - u_{zs})$ is an area and sector effect interacted with time due to the persistence of the initial conditions. This term disappears when t becomes large and the long run equilibrium is described by the term $y^{1*}_{zst} = \widetilde{\mathbf{x}}'_{zst}b + u_{zs}$. Equation (16) can then be rewritten as:

$$y^1_{zst} - y^{1*}_{zst} = \frac{(1-\rho)\eta_{zst}}{1-\rho L} \quad (19)$$

$$= \varepsilon_{zst}, \quad (20)$$

where ε_{zst} is a stationary noise.

Equation (17) describes the unconstrained model where both the levels and first differences of explanatory variables (including the lag of the dependent variable) affect the dependent variable (see equation 18). It yields therefore more complicated dynamics which might agree better with the usual view in empirical papers on economic growth. Indeed, many of them include both contemporary and past levels of explanatory variables in the regressions. Hence, evaluating the validity of the constraint $b_1 = \rho b$ leads to distinguish the different ways of modeling local employment growth in particular whether current and past local externality effects have to be introduced independently or in a constrained way.

In such an autoregressive panel data model, $y^1_{zs,t-1}$ depends on the area and sector effect, u_{zs}, if y^1_{zst} does. Moreover, given results of the previous section, variables $\widetilde{\mathbf{x}}_{zst}$ are likely to be correlated with the fixed effect u_{zs}. It leads us to estimate equation (17) by instrumental variable methods as in the previous section. Variables are supposed to be weakly exogenous that is to say, shocks in the equation of interest are not correlated with the past and present of explanatory variables. We use the two estimation methods that we have already presented. The equation is either estimated in first differences using instrument in levels (lagged once, twice or three times), or estimated in level using lagged first differences as instruments. Next, this is compared to the constrained estimation corresponding to equation (16).[4]

Compared to the static model, the dynamic information from the series of interest itself remains to be used (Hsiao, 1986). As estimates in table 5show that the order of the autocorrelation of the series Δy^1_{zst} is at least equal to 2 (and therefore the order of autocorrelation in the series y^1_{zst} at least equal to 1)), we have to use $y^1_{zs,t-3}$ or longer lags as instruments.

[4]To report on the strength of our instruments, table C.1 in appendix B reports results of instrumental regressions. They prove that the instruments we use are significant determinants of RHS variables in the equation of interest.

Estimation results are reported in the first two columns of table 6. Estimation results of the unconstrained equation (17) are quite similar and the value of the Sargan statistic associated to the instrumentation of differences by levels (column 2) does not lead to reject overidentifying restrictions. The estimate of the autoregressive coefficient is very precise[5]. Note also that the alternating signs of the coefficients of every explanatory variable and its lag agree well with the constrained specification (16).

That is why we also estimate the constrained model (i.e. under $b_1 = \rho b$). As this model is not linear, it is estimated by two stage non linear least squares (2SNLS) in first difference.[6] As in the unconstrained equation, the instruments are variables $y^1_{zs,t-3}$ and $\widetilde{\mathbf{x}}_{zs,t-j}$ for $j = 0, 1, 2$. Results are reported in the last column of table 6. Estimates are very precise even though standard errors of estimates should be corrected for biases due to the bilinear structure of the estimation method as well as for the presence of serial and spatial dependence.[7]

Estimated coefficients of explanatory variables in this constrained model are very similar to the estimates reported in table 5 for the static model. Therefore, these results can be interpreted as the long-run equilibrium between plant size and the explanatory variables, as written in equation (20).

At this point, it is informative to analyze residuals of the estimated equation and, among others, to determine the share of the variance explained by each group of variables, which is reported in table 7. The variance of the area and sector effects is smaller than in the univariate analysis reported in table B.1 of Appendix B. However, the decrease (20%) is not large, meaning that area and sector effects constant in time remain important determinants of the local firm size. Random shocks, η_{zst}, are well described as a moving average of order 1 which autocorrelation is equal to -0.140. This negative sign might reflect errors of measure, which is consistent with what could be expected. Indeed, measurement errors are plausible in the data since employment in an area and sector is reduced (increased) by a significant amount whenever a plant reduces (increases) the number of workers below (above) 20. Last we report an estimate of the variance of the "long-term" error, ε_{zst}, in the original series when the "short-term" error, η_{zst}, is assumed to be a MA(1) process. It is significantly larger that the estimated variance reported in table B.1 of Appendix B when covariates are omitted.

We also report in table 8 the contribution to the variance of the original series y^1_{zst} of the long-term target around which the series fluctuate, that is to say, $y^{1*}_{zst} = \widetilde{\mathbf{x}}'_{zst} b + u_{zs}$. The variance

[5] In very few estimations that we performed, did we find that higher-order lags of the dependent variables were significant. A first-order VAR process seems to be sufficient to describe these processes along with MA(1) random disturbances (see below).

[6] Note that the model is bilinear. Let z_{zst} a vector of instruments. Parameters ρ and b can then be estimated by using until convergence the following algorithm: Given a pair (ρ_n, b_n) obtained previously, (ρ_{n+1}, b_{n+1}) is estimated by regressing Δy_{zst} on variables $\Delta y_{zs,t-1}$ and $\Delta x_{zst} - \rho_n \Delta x_{zs,t-1}$ by 2SLS (instruments z_{zst}).

[7] In some experiments we ran, corrections for bilinearity and sectoral or serial dependence do not seem to matter. Spatial dependence matters more and makes Student statistics decrease by a maximum factor of 30%. These results are available upon request.

Table 6: Average plant size - Dynamic model

	(1)	(2)	(3)
Method	2SLS	2SLS	2SNLS
y^1_{zst}	L	Δ	$P\Delta$
Instruments	$\Delta y_{zs,t-2}$ $(\Delta\widetilde{\mathbf{x}}_{zs,t-j})_{j=0,1,2}$	$y_{zs,t-3}$ $(\widetilde{\mathbf{x}}_{zs,t-j})_{j=0,1,2}$	$y_{zs,t-3}$ $(\widetilde{\mathbf{x}}_{zs,t-j})_{j=0,1,2}$
y^2_{zst}	0.248 (20.6)	0.281 (4.9)	0.287 (20.1)
lL_{zt}	0.274 (13.9)	0.257 (1.6)	0.268 (15.3)
ls_{zt}	-0.126 (-3.2)	-0.093 (0.5)	-0.231 (8.7)
ld_{zt}	0.013 (-0.2)	0.166 (2.4)	0.136 (9.8)
$lcom_{zst}$	-0.505 (-38.3)	-0.497 (-11.9)	-0.576 (26.7)
c_{zst}	-0.053 (-6.4)	-0.002 (-0.1)	-0.063 (8.5)
$y^1_{zs,t-1}$	0.851 (30.7)	0.878 (16.3)	0.809 (20.1)
$y^2_{zs,t-1}$	-0.249 (25.1)	-0.257 (11.5)	-
$lL_{z,t-1}$	-0.233 (-11.1)	-0.237 (-4.2)	-
$ls_{z,t-1}$	0.221 (7.5)	0.239 (4.1)	-
$ld_{z,t-1}$	-0.156 (-7.1)	-0.109 (-4.4)	-
$lcom_{zs,t-1}$	0.504 (30.0)	0.529 (16.3)	-
$c_{zs,t-1}$	0.068 (8.2)	0.072 (4.9)	-
Sargan	0.025	0.839	0.521
$AC(1)$	0.061 ($<10^{-5}$)	-0.579 ($<10^{-5}$)	-0.576 ($<10^{-5}$)
$AC(2)$	0.194 ($<10^{-5}$)	0.084 ($<10^{-5}$)	0.082 ($<10^{-5}$)
$AC(3)$	0.194 ($<10^{-5}$)	-0.003 (0.57)	0.004 (0.53)
Obs	54664	54664	54664

Notes : **(a)** *Variables definition*: See table 2, note (b). **(b)** *L, Δ, and $P\Delta$* : Refer to variables in levels, Y_{zst}, first-differences, ΔY_{zst}, or in pseudo-differences. Pseudo-differences are build as $\Delta_\rho X_{zst} = X_{zst} - \rho X_{zs,t-1}$ where ρ is the coefficient of $y^1_{zs,t-1}$. We use the bilinear method explained in the text. **(c)** *Significativity*: Standard errors are not corrected for biases due to the bilinear nature of the estimation method nor for biases due to serial and spatial dependence. Experiments available upon request show that the correction factor is at most equal to 30%. Students statistics are reported in brackets. **(d)** *Sargan*: See table 5, note (d).

of the long-term target is equal to 85% of the total variance, which is high. The contribution of the variance of the area and industry effects to the variance of the target is also large (87%) while the variance of what is determined by explanatory variables contribute to it moderately (16%). These components do not sum to one as explanatory variables and area and sector effects are significantly and negatively correlated (-0.04). Average employment per plant is therefore moderately affected by the structural explanatory variables, $\widetilde{\mathbf{x}}_{zst} = (y^2_{zst}, \mathbf{x}_{zst})$, and much more by fixed effects, area and sector u_{zs}.

Table 7: Plant Size - Analysis of Variance

σ^2_u	σ^2_η	σ^2_ε	ρ_1	T	W
0.239	0.853	0.110	-0.140	0.01	71.72
(26.9)	(34.6)		(-12.8)	(0.92)	(0.028)

Notes : **(a)** *Notations*: σ^2_u is the variance of the area and sector effect, σ^2_η is the variance of the residual effect, ρ_1 its autocorrelation coefficient of order 1, T is the test statistic corresponding to a zero second order autocorrelation. W tests the validity of this representation of the series. We denote σ^2_ε the variance of $\varepsilon_{zst} = y_{zst} - \widetilde{\mathbf{x}}_{zst} + u_{zs}$. It is deduced from $\hat{\sigma}^2_\eta, \hat{\rho}_1$ and $\hat{\rho}$ using the expression: $\sigma^2_\varepsilon = V\left(\frac{1-\rho}{1-\rho L}\eta_{zst}\right) = \dfrac{1-\rho}{1+\rho}(1+2\rho\rho_1)\sigma^2_\eta$. **(b)** *Significativty*: Between brackets, Student statistics for coefficients and p-values for test statistics.

Table 8: Average plant size - Variance decomposition

	$\dfrac{Vy^{1*}_{zst}}{Vy_{zst}}$	$\dfrac{V(\widetilde{\mathbf{x}}'_{zst}\hat{b})}{Vy^*_{zst}}$	$\dfrac{V(u_{zs})}{Vy^*_{zst}}$	$Corr(\widetilde{\mathbf{x}}'_{zst}\hat{b}, u_{zs})$
y^1_{zst}	0.852	0.161	0.871	-0.042

Note: The long-term target is denoted $y^{1*}_{zst} = \widetilde{\mathbf{x}}'_{zst}\hat{b} + u_{zs}$ where coefficient $\hat{b}$ is taken from column 3 of table 6.

4.2 Number of Plants

In this section, we replicate the previous methodology to analyze the series of the number of plants in an area and sector. We start with the static specification and then turn to the constrained and unconstrained dynamic models.

The static model is written as:

$$y^2_{zst} = \check{\mathbf{x}}'_{zst}b + u_{zs} + \varepsilon_{zst}, \tag{21}$$

which differs from model (15) by the dependent variable, y^2_{zst} instead of y^1_{zst}, the inclusion of the lag of plant size, $y^1_{zs,t-1}$, among the explanatory variables and the absence of the local monopoly variable, c_{zst} that directly depends on the dependent variable. Let $\check{\mathbf{x}}_{zst}$ denote the new list of explanatory variables.

Estimated coefficients are reported in table 9. OLS estimates differ from all other estimates meaning that area and sector effects are still significantly correlated to other explanatory variables at a lesser degree than for plant size, however. The divergence between signs across different columns is also less noticeable than for plant size. Moreover, whereas the Sargan statistic of the estimation reported in column 4 indicates that overidentifying restrictions are not rejected, it is not the case of estimates reported in columns 5 and 6 (estimations in first differences). Differences are small between estimates, however.

Table 9: Number of plants - Static model

	(1)	(2)	(3)	(4)	(5)	(6)
Method	OLS	Within	OLS	2SLS	2SLS	2SLS
y^2_{zst}	L		Δ	L	Δ	Δ
Instruments	-	-	-	$(\Delta\mathbf{x}_{zs,t-j})_{j=1,2,3}$	$(\mathbf{x}_{zs,t-j})_{j=2,3,4}$	$(\mathbf{x}_{zs,t-j})_{j=1,2,3}$
$y^1_{zs,t-1}$	0.222 (143.1)	0.033 (14.2)	−0.040 (−16.8)	−0.012 (−0.5)	−0.114 (−3.2)	−0.012 (−1.5)
lL_{zt}	0.161 (76.3)	0.263 (35.9)	0.244 (33.3)	0.202 (3.9)	0.225 (2.1)	0.233 (8.4)
ls_{zt}	−0.247 (−24.9)	−0.111 (−8.9)	−0.097 (−7.8)	−0.196 (−1.8)	−0.205 (−1.4)	−0.162 (−3.5)
ld_{zs}	0.071 (22.2)	0.117 (19.6)	0.103 (16.2)	0.217 (2.8)	0.024 (0.4)	0.076 (3.4)
$lcom_{zst}$	1.04 (620.7)	0.853 (363.0)	0.850 (356.9)	0.937 (55.3)	0.697 (13.1)	0.815 (72.5)
R^2	92.9%	68.8%	67.7%	-	-	-
Sargan	-	-	-	0.80	0.002	5.10^{-5}
AC(1)	0.825 ($<10^{-4}$)	0.974 ($<10^{-4}$)	−0.253 ($<10^{-4}$)	0.908 ($<10^{-4}$)	−0.215 ($<10^{-4}$)	−0.244 ($<10^{-4}$)
AC(2)	0.766 ($<10^{-4}$)	0.962 ($<10^{-4}$)	−0.045 ($<10^{-4}$)	0.860 ($<10^{-4}$)	−0.079 ($<10^{-4}$)	−0.058 ($<10^{-4}$)
AC(3)	0.705 ($<10^{-4}$)	0.951 ($<10^{-4}$)	−0.034 (10^{-4})	0.818 ($<10^{-4}$)	−0.044 ($<10^{-4}$)	−0.038 ($<10^{-4}$)
Obs	72780	72780	63523	37826	37826	46099

Notes : **(a)** *Variables definition*: See table 2, note (b). **(b)** *L and* Δ: See table 5, note (d). **(c)** *Significativity*: See table 5, note (c). **(d)** *Sargan*: See table 5, note (d).

Estimation results for the two dynamic specifications, translated from equations (16) and (17) for plant creation, are reported in table 10. The estimation of the unconstrained dynamic model yields similar results whatever the estimation method is. The signs of the coefficients of each variable and its lag are alternating again, which confers credibility to the constrained specification. Estimates in the pseudo-differenced equation are very precise. Sargan statistics indicate that overidentifying restrictions cannot be rejected at a reasonable level of significance. As for plant size, the order of autocorrelation is equal to 2 in first differences, which corresponds to a moving average of order 1 when the dependent variable is in level. Estimates in the dynamic model do not differ much from those in the static model. Specification diagnostics are better however in the dynamic version. All of these allow us to conclude that the dynamics of plant creation, as the one of plant size, is correctly captured by a model assuming random oscillations around a time dependent long run target corresponding to the static model.

Table 10: Number of plants - Dynamic model

	(1)	(2)	(3)
Method	2SLS	2SLS	2SNLS
y^2_{zst}	L	Δ	$P\Delta$
Instruments	$\Delta y_{zs,t-2}$ $(\Delta\widetilde{\mathbf{x}}_{zs,t-j})_{j=0,1,2}$	$y_{zs,t-3}$ $(\widetilde{\mathbf{x}}_{zs,t-j})_{j=0,1,2}$	$y_{zs,t-3}$ $(\widetilde{\mathbf{x}}_{zs,t-j})_{j=0,1,2}$
$y^1_{zs,t-1}$	-0.047 (-7.2)	-0.041 (-1.7)	-0.042 (-10.1)
lL_{zt}	0.246 (20.5)	0.191 (1.9)	0.243 (18.5)
ls_{zt}	-0.110 (-3.4)	-0.107 (-0.8)	-0.108 (-5.3)
ld_{zt}	0.099 (2.9)	0.082 (1.5)	0.096 (9.3)
$lcom_{zst}$	0.848 (168.2)	0.818 (32.1)	0.831 (37.1)
$y^2_{zs,t-1}$	0.787 (34.5)	0.829 (15.6)	0.710 (27.4)
$y^1_{zs,t-2}$	0.023 (4.6)	0.036 (5.1)	-
$lL_{z,t-1}$	-0.189 (-11.4)	-0.219 (-6.4)	-
$ls_{z,t-1}$	0.061 (2.9)	0.092 (2.4)	-
$ld_{z,t-1}$	-0.067 (-4.7)	-0.086 (-4.5)	-
$lcom_{zs,t-1}$	-0.661 (-32.7)	-0.692 (-16.1)	-
Sargan	0.065	0.300	0.122
$AC(1)$	0.075 $(<10^{-5})$	-0.571 $(<10^{-5})$	-0.561 $(<10^{-5})$
$AC(2)$	0.204 $(<10^{-5})$	0.062 $(<10^{-5})$	0.058 $(<10^{-5})$
$AC(3)$	0.203 $(<10^{-5})$	-0.001 (0.91)	-0.003 (0.71)
Obs	46099	54664	54664

Notes : **(a)** *Variables definition*: See table 2, note (b). **(b)** *L, Δ, and $P\Delta$* : See table 6, note (d). **(c)** *Significativity*: See table 6, note (c). **(d)** *Sargan*: See table 5, note (d).

Table 11: Number of plants - Analysis of variance

σ^2_u	σ^2_η	σ^2_ε	ρ_1	T	W
0.090 (33.4)	0.173 (36.0)	0.027	-0.123 (-11.3)	0.90 (0.34)	74.78 (4.10^{-5})

Notes : **(a)** *Notations*: See table 7, note (b). **(b)** *Significativity*: See table 7, note (c).

Table 12: Number of plants - Variance decomposition

	$\frac{Vy^{2*}_{zst}}{Vy_{zst}}$	$\frac{V(\mathbf{x}'_{zst}\hat{b})}{Vy^*_{zst}}$	$\frac{V(u_{zs})}{Vy^*_{zst}}$	$Corr(\mathbf{x}'_{zst}\hat{b}, u_{zs})$
y^2_{zst}	0.744	0.965	0.004	0.263

Note: The long-term target is denoted $y^{1*}_{zst} = \widetilde{\mathbf{x}}'_{zst}\hat{b} + u_{zs}$ where coefficient $\hat{b}$ is taken from column 3 of table 10.

Finally, we can analyze residuals as in the previous section (tables 11 and 12). Autocorrelation estimates confirm results of table 10. Residuals η_{zst} are a moving average of order 1 and the coefficient of autocorrelation is equal to -0.123, which might reflect measurement errors as before. The variance of the area and sector effect is much smaller (1/10) than in the original series analyzed in table B.1. The variance of η_{zst} is also much smaller. The model is thus better suited for explaining (the logarithm of) the number of plants than plant size. It is confirmed by what we report in table 12. The variance of the long-term target is equal to 74% of the variance of the original series and the long-term variance is mainly composed of the variance of what is determined by explanatory variables (96%). Area and sector effects contribute very little to this variance (0.4%) and the correlation between explanatory variables and area and sector effects is very significant (0.26). The number of plants is therefore well explained by the model.

Finally, as complementary results, what can be obtained from a disaggregated analysis by sectors (manufacturing, trade, and services) is reported in tables D.1 and D.2 in appendix D. They show that measures of persistence and the effect of explanatory variables on both dependent variables are quite similar across sectors even if some small differences underlined in the next section may be observed. The significance of estimates also slightly differ across sectors but it might only reflect small sampling errors in the samples concerning trade and services.

4.3 Interpretations

In this section we put into perspective the estimation results of the dynamic models. Do not forget that some caution should nevertheless be exercised when interpreting our results since creation and destruction of employment in small plants, where less than 20 employees work, should also be taken into account (Davis, Haltiwanger and Schuh, 1996) in order to obtain a complete view of the local employment and firm creation dynamics.

We start with two important general conclusions that stem from the specification that is finally selected, a stationary noise around a long run target corresponding to the static model. First, this implies that the dynamic components of both firm size and firm creation are secondary compared to the effects of current local conditions. Not only the static model alone fairly well explains both dependent variables, but the constrained dynamic model corresponding to an ARMA(1,1) is also accepted. In other words, our data seems more to sustain a static setting, such as those recently developed in economic geography based on static trade models under imperfect competition, that is to say static externalities, rather than dynamic externalities and growth models. By contrast, Henderson [1997] finds for the U.S. much more persistence in the local employment dynamics and the impact of the local economic structure. Second, we also show that regarding plant size, area and sector fixed effects significantly and strongly matter in the determination of the long-run target. Therefore, plant size seems to be poorly affected by local externalities due to interactions between agents. By contrast, local fixed endowments such as physical geography (proximity to a border, to an ocean, presence of mountains, of mines etc), local public goods or permanent comparative advantages seems to play an important role. The number of plants, on the other hand, is much more influenced by the local economic structure and therefore by local externalities due to the interactions between agents, even if these externalities are shown to be simultaneously more correlated with the fixed area and sector effects.

Dynamic effects are not zero, however. The persistence of shocks is measured by the autoregressive coefficient, ρ. It gives evidence of a mean reversion process. Estimates of this coefficient for both series lie between 0.75 and 0.81 for plant size and between 0.50 and 0.73 for plant number. When the dependent variable is plant size, there is little difference in persistence between sectors (table D.1). By contrast, persistence in the number of plants is more variable across sectors (table D.2). Trade and services differ from manufacturing by a weaker persistence in their creation/destruction of plants. As a conclusion, Marshall Arrow Romer externalities are not observed, which would be the case if the autoregressive coefficient would be larger than one, but some inertia in both plant size and number of plants is present, even more as regards the former. This sustains the idea that once an industry in a given area present larger plant size and plant number, this remains the case for a number of periods but not indefinitely. Moreover, since the constrained model can be read in first differences, comparable statements can be done for growth rates. High external or internal growth rates at a given period imply high growth rates in the future, at least in the short run, even more for plant size and for manufacturing.

The higher persistence of firm size and the larger share of variance explained by the area and sector fixed effect in this case have an important policy implication. Actually, economic policies affecting the local economic structure would be more efficient on the plant creation, that is external growth, than those trying to influence the growth of existing plants. Moreover, they would be also more effective in trade and services than in manufacturing.

Contrary to previous studies, we are able to distinguish this short-run dynamic effects from the impact of the local economic structure conditional on the presence of area and sector effects. We therefore turn now to the respective impact of each variable on which economic policies could be targeted. Let us start by the impact on local growth of the global size of the local area (lL_{zt}). We find that in larger areas, cities as opposed to more rural areas for instance, both internal and external growth are stronger. It is consistent with the ongoing increase in urban concentration observed in the U.S. by Black and Henderson [1998] even if these authors also underline recent examples of small but rapidly growing cities. Combes [2000] also concludes for France that growth is more important in large cities as regards service sectors but the reverse happens in manufacturing.

Hence, the data exhibit global agglomeration economies. Larger areas where both technological spillovers might be stronger and where final and intermediate good markets are larger have larger industry plant size growth as well as stronger industry plant creation. This last point is in particular consistent with the idea of "nursery" cities developed by Duranton and Puga [2001]. Cities, where ideas and knowledge are concentrated, would be the most favorable places for creating and innovating at the first stage of the product life cycle, activity moving towards less dense areas at later stages. It appears to be even more valid in France as regards service activities which makes sense if these products are more innovation intensive and frequently renewed.

The global size of the local economy is sometimes considered as part of the so called urbanization economies. Nevertheless, this latter term has recently been meant to describe the impact of local sectoral diversity (ld_{zt}) on local growth. To describe the local industrial composition, we also included the number of active sectors (ls_{zt}) as an extra explanatory variable next to the diversity index. Both variables are always significant in all sectors. Diversity has a positive effect on plant size and as on the number of plants. In contrast, the number of operating sectors in the area (ls_{zt}) has a negative effect in both cases.

As a consequence, the message regarding the role of the sectoral composition is refined. The most favorable local industrial structure would consist in a small number of sectors but of roughly the same size, which would maximize both the industry internal and external growth. In terms of agglomeration forces, it is consistent with the idea that cost and demand linkages extend similarly to all intermediate inputs in a given industry, even if the number of these inputs is not necessary large. As regards pure local externalities, technological spillovers might be cross-sectoral but would not extend to all sectors. They would be maximized inside fairly small but balanced sub-groups of similar sized sectors.

Ceteris paribus, the effect of the number of plants (y^2_{zst}) on plant size (y^1_{zst}) is positive. In other terms, employment in an area and sector grows more quickly than the number of plants. This can be interpreted in terms of market based and pure local externalities. Using the imperfect competition set up developed in section 2 allow us to be more precise. Actually, the result can be reconciled with equation (9) if the supply elasticity of labor is sufficiently large, which seems natural enough in such a well integrated economy as the French economy. The direct negative effect of the number of plants on firm size is lower in this case than the positive indirect one due to the productivity gains that decrease price, which next increases demand when the demand elasticity is greater than 1. The effect of the lagged plant size (y^1_{zst}) on the number of plants (y^2_{zst}) is small but negative. It might be due to the presence of large recurring fixed costs which might partly deter entry into the market. Second, local competition between plants within sectors ($lcom_{zst}$) has a strong and negative effect on plant size and has a strong and positive effect on the number of plants. The more concentrated the plants are, the more employment is growing within plants though the smaller plant creation is. One might argue that this variable is strongly correlated with the number of plants in an area and industry. We ran again the regression of plant creation omitting the competition variable, $lcom_{zst}$, and we saw that the magnitude of the effects of other explanatory variables are affected though not their signs. Finally, the effect of the variable that denotes a monopoly situation (c_{zst}) goes in the reverse direction as regards plant size. [8] In other words, competition seems to have non-linear effects on the growth of areas and sectors.

Even if more complex, these results about the effects of competition shed new light on what is found in the literature and have also some important policy implications. In Glaeser *et al.* [1992], competition is proxied by the (inverse) size of plant and results can be hardly compared to our approach. In Combes [2000] only one competition variable, $lcom_{zst}$, is considered. It has a quasi systematic negative effect, except in a few service sectors. Again, the inclusion of different competition variables here, and the fact that they are all significant (except the monopoly situation in services), allows a more precise verdict. Internal growth is maximized when the number of local plants is large, when they are definitely not in a situation of local monopoly and when, simultaneously, the size of plants is uneven. One can think about a situation in which some large leader firms impulse the dynamics of a large number of smaller plants. The size of the first ones would induce benefits due to economies of scale while the other ones could for instance gain from cost and demand linkages. It is also consistent with the idea of large plants doing research and development that benefits to all other plants, a structure that would maximize technological spillovers and knowledge diffusion. The counterpart would be that, once controlled by the mean reversion effect (plant creation is stronger in areas where the number of plants is low), the existence of plants of even size is more favorable to the entry of new plants.

5 Conclusion

We analyze yearly data extracted from an employment survey of plants to study the impact in France of the local economic structure on local 1984-1993 growth of the industry plant size and number of plants simultaneously. The short-run dynamics is carefully studied and we control for fixed effects and endogeneity, which prove to be important devices. The optimal local economic structure in terms of total size, industrial composition, and plant size distribution is precisely characterized. This is interpreted in the light of the imperfect competition model we present and more generally of urban economics and economic geography. Economic policy implications

[8] As already said, this variable has obviously been omitted from the equation explaining the number of plants.

regarding both internal and external growth are derived.

Replication of studies such as ours would make possible a precise characterization of the different forms of local development that are undertaken in each of the countries of the European Union and to a factual analysis of the efficiency of creating employment according to these references. Besides, we are not able to state which kind of agglomeration externalities, technological or market-based, are more important for local growth, a question on which more research efforts should certainly be put. Another line of research would also consist in evaluating the spatial extent to which the local structure acts on local growth. We assume here that it is restricted to the local area itself, but Desmet and Fafchamps [2001] for instance propose a methodology that could be mixed to ours to evaluate the distance at which agglomeration forces work. Finally, we did not analyze separately the influence of local infrastructure and more generally of public goods available at the local level, which are simply controled simultaneously with other endowment effects by fixed effects. Research on these issues is certainly high in the agenda.

References

[1] **J. Abowd and D. Card**, On the Covariance Structure of Earnings and Hours Changes, *Econometrica*, **57**, 411-445 (1989).

[2] **J. Altonji and L. Segal**, Small-Sample Bias in GMM Estimation of Covariance Structures, *Journal of Business and Economic Statistics*, **14**, 353-366 (1996).

[3] **R. Barro, R. and X. Sala-I-Martin**. *Economic Growth,* MIT Press (1995).

[4] **D. Black and V. Henderson**, Spatial Evolution of Population and Industry in the U.S.A, *American Economic Review*, **89**, 321-327 (1999).

[5] **R. Blundell and S. Bond**, Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, **87**, 115-143 (1998).

[6] **P.-P. Combes**, Economic Structure and Local Growth: France, 1984-1993, *Journal of Urban Economics*, **47**, 329-355 (2000).

[7] **P.-P. Combes and M. Lafourcade**, Transportation Costs Decline and Regional Inequalities: Evidence from From, *CEPR Discussion paper*, **2894** (2001).

[8] **P.-P. Combes and H.G. Overman**, The Spatial Distribution of Economic Activities in the European Union, in *Handbook of Urban and Regional Economics, vol. 4*, V. Henderson and J.-F. Thisse (eds.), Elsevier-North Holland, Amsterdam (2004).

[9] **S. Davis, J. Haltiwanger and S. Schuh**, Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts, *Small Business Economics*, **8**, 297-315 (1996).

[10] **K. Desmet and M. Fafchamps**, The Changing Spatial Distribution of Economic Activity across U.S. Counties, *Oxford University mimeo* (2001).

[11] **G. Duranton and D. Puga**, Nursery Cities: Urban diversity, process innovation, and the life cycle of products, *American Economic Review*, **91**, 1454-1477 (2001).

[12] **D. Encaoua and A. Jacquemin**, Degree of Monopoly, Indices of Concentration and Threat of Entry, *International Economic Review*, **21**, 87-105 (1980).

[13] **E. L. Glaeser, H. Kallal, J. Sheinkman and A. Schleifer**, Growth in Cities, *Journal of Political Economy*, **100**, 1126-1152 (1992).

[14] **C. Gouriéroux and A. Monfort**, *Statistics and Econometric Models*, Cambridge University Press, Cambridge (1999).

[15] **J.A. Hausman and T. Taylor,** Panel Data And Unobservable Fixed Effects, *Econometrica*, **49**, 1377-98 (1981).

[16] **J. V. Henderson**, Externalities and industrial development, *Journal of Urban Economics*, **42**, 449-470 (1997).

[17] **J. V. Henderson, A. Kuncoro and M. Turner**, Industrial Development in Cities, *Journal of Political Economy*, **103**, 1067-1090 (1995).

[18] **C. Hsiao**, *Analysis of Panel Data*, Cambridge University Press, Cambridge (1986).

[19] **R. Ziliak**, Efficient Estimation with Panel Data when Instruments are Predetermined: An Empirical Comparison of Moment Condition Estimators, *Journal of Business and Economic Statistics* , **15**, 419-31 (1997).

Appendices

A Theoretical Predictions

A.1 Perfect Competition

Differentiating the production function (1) yields:

$$\frac{dY_z}{Y_z} = \frac{dA_z}{A_z} + \alpha\frac{dL_z}{L_z} + (1-\alpha)\frac{dK_z}{K_z}. \tag{A.1}$$

As the price is equal to the marginal cost, we also have:

$$\frac{dp_z}{p_z} = \alpha\frac{dw_z}{w_z} - \frac{dA_z}{A_z}. \tag{A.2}$$

Equalization of labor productivity to real wage is given by:

$$\frac{w_z}{p_z} = \left(\frac{\alpha}{1-\alpha}\right)^{(1-\alpha)} A_z \left(L_z\right)^{\alpha-1} \left(K_z\right)^{1-\alpha}, \tag{A.3}$$

and differentiation yields:

$$\frac{dw_z}{w_z} - \frac{dp_z}{p_z} = \frac{dA_z}{A_z} - (1-\alpha)\frac{dL_z}{L_z} + (1-\alpha)\frac{dK_z}{K_z}. \tag{A.4}$$

Given the definitions of elasticities (equation 3), we then have six linear equations in six unknowns: $\frac{dY_z}{Y_z}$, $\frac{dL_z}{L_z}$, $\frac{dK_z}{K_z}$, $\frac{dp_z}{p_z}$, $\frac{dw_z}{w_z}$, $\frac{dr_z}{r_z}$ for $\frac{dA_z}{A_z}$ given. We therefore obtain equation (4) reported in the text.

A.2 Cournot Competition: Short Run

Profit per firm is given by:

$$\pi_z = (p_z - c_z)\, q_z - f, \tag{A.5}$$

where

$$c_z = \frac{(w_z)^{\alpha}\, r^{1-\alpha}}{\alpha^{\alpha}\,(1-\alpha)^{(1-\alpha)}\, A_z}$$

is the marginal cost and q_z is the individual production.

The first-order condition is:

$$(p_z - c_z) + \frac{\partial p_z}{\partial q_z} q_z = 0,$$

and the Cournot assumption yields:

$$\frac{\partial p_z}{\partial q_z} = -\frac{1}{\sigma}\frac{p_z}{Y_z},$$

which results in

$$\frac{(p_z - c_z)}{p_z} = \frac{1}{\sigma}\frac{q_z}{Y_z},$$

and therefore

$$\frac{(p_z - c_z)}{p_z} = \frac{1}{\sigma N_z}. \tag{A.6}$$

The equilibrium price is given by equation (5). Differentiation yields:

$$\frac{dp_z}{p_z} = \alpha\frac{dw_z}{w_z} - \frac{dA_z}{A_z} + \frac{dN_z}{N_z} - \frac{\sigma dN_z}{\sigma N_z - 1},$$

which implies:

$$\frac{dp_z}{p_z} = \alpha\frac{dw_z}{w_z} - \frac{dA_z}{A_z} - \frac{1}{\sigma N_z - 1}\frac{dN_z}{N_z}. \tag{A.7}$$

As equation (A.7) replaces equation (A.2), there are again six equations for six unknowns as N_z and $\frac{dN_z}{N_z}$ are exogenous in short-run. We therefore obtain equation (A.8) as identical to equation (6) in the text:

$$\frac{dL_z}{L_z} = \frac{\varepsilon\,(\sigma-1)}{1-\alpha+\varepsilon+\alpha\sigma}\left(\frac{dA_z}{A_z} + \frac{1}{\sigma N_z - 1}\frac{dN_z}{N_z}\right). \tag{A.8}$$

Using

$$\frac{dL_z}{L_z} = \frac{dl_z}{l_z} + \frac{dN_z}{N_z},$$

we obtain:

$$\frac{dl_z}{l_z} = \frac{\varepsilon(\sigma - 1)}{1 - \alpha + \varepsilon + \alpha\sigma}\left(\frac{dA_z}{A_z} + \frac{1}{\sigma N_z - 1}\frac{dN_z}{N_z}\right) - \frac{dN_z}{N_z}. \tag{A.9}$$

A.3 Cournot Competition: Long Run

The first order condition (A.6) yields:

$$\pi_z = \frac{p_z}{\sigma N_z} q_z - f = \frac{p_z}{\sigma N_z^2} Y_z - f = 0.$$

By differentiation,

$$\frac{dp_z}{p_z} + \frac{dY_z}{Y_z} - 2\frac{dN_z}{N_z} = 0.$$

There are now seven equations for seven unknowns, $\frac{dN_z}{N_z}$ being the seventh. When equation (A.9) is used, there are eight equations for eight unknowns, $\frac{dl_z}{l_z}$ being the eighth.

Long-run variations are therefore given by equations (9), (10), and (11) as in the text. For instance:

$$\frac{dl_z}{l_z} = \frac{(\varepsilon - 1)(\sigma - 1)(\sigma N_z - 1)}{2\alpha(\sigma - 1)(\sigma N_z - 1) + (1 + \varepsilon)(\sigma(N_z - 1) + \sigma N_z - 1)}\frac{dA_z}{A_z}.$$

The denominator are the same for $\frac{dL_z}{L_z}$ and $\frac{dN_z}{N_z}$, positive if $\sigma > 1$ and $N_z > 1$. $\frac{dL_z}{L_z}$ and $\frac{dN_z}{N_z}$ numerators are positive under the same conditions. $\frac{dl_z}{l_z}$ also needs $\varepsilon > 1$ to be positive.

B Univariate Analysis

As every variable varies in three dimensions z, s, and t and as we demeaned variables with respect to sector and period effects, we proceed by a descriptive analysis of the covariance structure of each variable setting forth their dependence on area and sector effects and the interactions of those with time. For any component, say X_{zst}, of the vector of endogenous variables $\mathbf{y}_{zst}$ or of determinants $\mathbf{x}_{zst}$, we analyze its cross-section and dynamic characteristics by writing:

$$X_{zst} = u_{zs} + \delta_t v_{zs} + \varepsilon_{zst}, \quad \varepsilon_{zst} \sim MA(q), \tag{B.1}$$

where random shocks u_{zs} and ε_{zst} are uncorrelated and where ε_{zst} is a moving average of an *a priori* unknown order equal to q. This representation, though by no means the most general, is well suited to summarize cross-section correlations (that is between pairs (z, s) through random variables such as u_{zs} and v_{zs}) and how these cross-section correlations vary with time (through parameters δ_t), as well as the serial dependence of each history specific to a given pair (z, s) through the random shock ε_{zst} (see Hsiao, 1986).

The different components of model (B.1) are estimated by minimizing the distance between the unrestricted variance-covariance matrix of X_{zst} across time, which entries are $E(X_{zst}X_{zst'})$, and the variance-covariance matrix restricted by different specification of equation (B.1) (Abowd and Card, 1989).[9] Results are reported in table B.1. The first five columns report estimates of the variance of the area and sector effect u_{zs} and of the variance of the specific shock ε_{zst} as well as its first three serial dependence coefficients if the following constrained specification :

$$X_{zst} = u_{zs} + \varepsilon_{zst}, \quad \varepsilon_{zst} \sim MA(3), \tag{B.2}$$

is verified ($\delta_t = 0$). In other columns we report results of various specification tests. Using statistics W, specification (B.2) against a unrestricted alternative can be tested. Using T_1, we test that the fourth-order autocorrelation is equal to zero, that is to say specification (B.2) against :

$$X_{zst} = u_{zs} + \varepsilon_{zst}, \quad \varepsilon_{zst} \sim MA(4). \tag{B.3}$$

Last, using statistics T_2, we test that cross-section correlations are stable, that is to say specification (B.2) against the alternative:

$$X_{zst} = u_{zs} + \delta_t v_{zs} + \varepsilon_{zst}, \quad \varepsilon_{zst} \sim MA(3). \tag{B.4}$$

First, specification (B.2) is always rejected against an unrestricted alternative (column W). According to columns T_1 and T_2, it mainly stems from the instability of cross-section correlations and not from dynamics of higher order. The codependence between series in various area and sector (z, s) might not be stable across time. The hypothesis that the fourth-order autocorrelation is equal to zero is however rejected for three series out of 7 and in particular those which consist in aggregates at the geographical level.

The respective variances of area and sector effects u_{zs} and idiosyncratic shocks ε_{zst} differ by an order of magnitude. The variance of the former is between 2 and 350 times larger than the variance of the latter. Variability is therefore much larger in the spatial and sectoral dimension than in the time dimension. It is particularly true in the case of local aggregates such as (in logarithms) total employment and the number of active sectors in an area. It can be noted however that the time dimension is the poorest (10) among the three dimensions (sectors, 36, areas, 341).

This decomposition can be repeated for growth rates, i.e. the first differences of the series. After differentiating the model (B.1) becomes:

$$\Delta X_{zst} = (\Delta\delta_t)v_{zs} + \Delta\varepsilon_{zst}, \quad \varepsilon_{zst} \sim MA(q), \tag{B.5}$$

If coefficients δ_t are a linear function of time the difference $\Delta\delta_t$ is constant. The estimation of an error component structure such as model (B.2) for first differenced series (which is not

[9] The weighting matrix in the minimum distance procedure is the optimal weighting matrix computed from the sample. Small sample biases are neglected (see Altonji and Segal, 1996).

reported here) leads however to reject such a case. Nevertheless, the absence of area and sector effects is rejected for the series in first-differences (see table B.2) because correlations between the dynamics of the series for each pair (z, s) vary with time and because $\Delta\delta_t$ is not constant (except for the first series y^1_{zst} where all $\Delta\delta_t$ are equal to zero).

Table B.1: Univariate statistical analysis of each series (in levels)

	σ^2_u	σ^2_ε	ρ_1	ρ_2	ρ_3	T_1	T_2	W
y^1_{zst}	0.28	0.035	0.48	0.26	0.10	1.96	136.7	557.06
	(30.9)	(28.6)	(41.1)	(24.0)	(14.6)	(0.16)	($<10^{-5}$)	($<10^{-5}$)
y^2_{zst}	0.85	0.046	0.51	0.28	0.11	7.92	55.7	872.06
	(37.3)	(41.2)	(72.9)	(35.1)	(18.3)	(0.005)	($<10^{-5}$)	($<10^{-5}$)
lL_{zt}	0.73	0.0024	0.47	0.25	0.14	4.94	278.4	1679.7
	(34.0)	(34.0)	(50.0)	(28.9)	(22.0)	(0.03)	($<10^{-5}$)	($<10^{-5}$)
ls_{zt}	0.033	0.0012	0.52	0.31	0.13	23.0	427.0	1364.1
	(34.5)	(37.9)	(76.0)	(42.9)	(24.0)	($<10^{-5}$)	($<10^{-5}$)	($<10^{-5}$)
ld_{zt}	0.087	0.0025	0.45	0.21	0.10	32.6	404.9	1890.7
	(25.6)	(29.3)	(42.0)	(22.4)	(13.9)	($<10^{-5}$)	($<10^{-5}$)	($<10^{-5}$)
$lcom_{zst}$	0.50	0.041	0.52	0.28	0.12	0.40	429.4	941.86
	(35.5)	(40.6)	(68.3)	(36.0)	(18.8)	(0.53)	($<10^{-5}$)	($<10^{-5}$)
c_{zst}	0.07	0.031	0.46	0.23	0.10	1.86	136.7	347.4
	(39.0)	(29.1)	(42.5)	(20.9)	(11.4)	(0.17)	($<10^{-5}$)	($<10^{-5}$)

Notes : **(a)** *Variables definition*: See table 2, note (b). **(b)** Each series is decomposed into (*) $X_{zst} = u_{zs} + \varepsilon_{zst}$ where u_{zs} and ε_{zst} are uncorrelated. σ^2_u is the variance of $u_{zs}, \sigma^2_\varepsilon$ is the variance of ε_{zst}, ρ_i are autocorrelations of order 1, 2 and 3 of ε_{zst}. Student statistics are reported between brackets. **(c)** Column W reports the test statistics of specification (*). If (*) is true, W is distributed as a chi-square with 51 degrees of freedom. Statistics T_1 is used to test that the fourth-order autocorrelation is equal to zero. Under the null, it is distributed as a chi-square with one degree of freedom. Statistics T_2 is used to test that parameters δ_t in the specification (**) $X_{zst} = u_{zs} + \delta_t v_{zs} + \varepsilon_{zst}$ (u_{zs}, v_{zs}, ε_{zst} uncorrelated) are equal to zero. Under the null, it is distributed as a chi-square with nine degrees of freedom. p-values are given between brackets. **(d)** The data are balanced for each series.

Table B.2: Univariate statistical analysis of first-differenced series

	T_1	T_2	W
Δy^1_{zst}	0.74	12.4	81.44
	(0.39)	(0.13)	(2.10^{-4})
Δy^2_{zst}	0.01	27.1	202.92
	(0.92)	(6.10^{-4})	($<10^{-5}$)

Notes : **(a)** *Variables definition*: See table 2, note (b). **(b)** Each series ΔX_{zst} is supposed to be autocorrelated of order 3. Area and sector effects are not allowed for since their presence is always rejected. **(c)** Statistics W is used to test the hypothesis that $\Delta X_{zst} = \varepsilon_{zst}$, $\varepsilon_{zst} \sim MA(3)$. It is distributed as a chi-square with 42 degrees of freedom under the null hypothesis. Statistics T_1 is used to test that the fourth-order autocorrelation is equal to zero. Under the null, it is distributed as a chi-square with one degree of freedom. Statistics T_2 is used to test that parameters δ_t in the model $\Delta X_{zst} = \delta_t v_{zs} + \varepsilon_{zst}$ (u_{zs} , ε_{zst} uncorrelated) are equal to zero. Under the null, it is distributed as a chi-square with eight degrees of freedom. p-values are given between brackets. **(d)** The data are balanced for each series.

Concluding, these results suggest a specification such that:

$$X_{zst} = \alpha X_{zs,t-1} + e_{zs} + \eta_{zst}. \tag{B.6}$$

Indeed, expanding the series we get:

$$X_{zst} = \alpha^t X_{zs0} + \frac{1-\alpha^t}{1-\alpha} e_{zs} + \eta_{zst} + \alpha\eta_{zst-1} + \alpha^2\eta_{zs,t-2} + ... + \alpha^{t-1}\eta_{zs1}.$$

Terms such as $\delta_t v_{zs}$ in the previous models could be compatible with the geometric effect of initial conditions ($\alpha^t(X_{zs0} - e_{zs}/(1-\alpha))$, and the presence of an area and sector effect, $e_{zs}/(1-\alpha)$. The smoothly decreasing and positive autocorrelations of shocks ε_{zst} (ρ_1, ρ_2 and ρ_3 in table 3) are also compatible with the autocorrelation structure induced by such a specification of the series.

C Instrumental regressions

Table C.1: Instrumental regressions

(Obs 54642)	Δy^1_{zst}	Δy^2_{zst}	$\Delta lcom_{zst}$	Δc_{zst}	Δld_{zs}	ΔlL_{zt}	Δls_{zs}
$y^1_{zs,t-1}$	−0.270 (−58.6)	–	–	–	–	–	–
$y^1_{zs,t-2}$	0.150 (27.3)	0.024 (−4.0)	–	−0.014 (−2.7)	0.004 (1.8)	–	–
$y^1_{zs,t-3}$	0.058 (13.4)	0.011 (2.3)	–	–	−0.003 (−1.9)	−.003 (1.7)	–
$y^2_{zs,t-1}$	−0.040 (−5.1)	−0.262 (−30.9)	0.078 (9.6)	−.042 (−5.8)	–	0.006 (2.7)	–
$y^2_{zs,t-2}$	0.018 (2.0)	0.130 (12.9)	−0.018 (1.9)	0.021 (2.5)	–	–	–
$y^2_{zs,t-3}$	0.038 (5.0)	0.093 (11.1)	–	0.015 (2.1)	–	−.007 (−2.2)	–
$lL_{z,t-1}$	–	−0.024 (−1.8)	−0.026 (2.0)	0.023 (2.0)	−0.025 (−4.9)	−0.340 (−75.3)	−0.006 (−2.4)
$lL_{z,t-2}$	–	0.076 (5.0)	0.045 (3.1)	−0.041 (−3.1)	–	0.259 (50.5)	0.013 (4.3)
$lL_{z,t-3}$	–	−0.031 (−2.7)	–	–	0.022 (5.0)	0.081 (21.2)	0.006 (−2.7)
$ls_{z,t-1}$	0.086 (3.9)	–	–	–	0.141 (15.7)	–	−0.317 (−68.0)
$ls_{z,t-2}$	−0.060 (−2.3)	–	0.052 (1.9)	–	−0.048 (−4.5)	0.046 (4.9)	0.200 (36.3)
$ls_{z,t-3}$	–	–	−0.048 (−2.2)	–	−0.024 (−2.7)	−0.065 (−8.6)	0.055 (12.3)
$ld_{z,t-1}$	–	–	–	–	−0.321 (−70.4)	−0.017 (−4.4)	0.007 (2.9)
$ld_{z,t-2}$	–	–	–	–	0.186 (34.2)	–	0.011 (4.0)
$ld_{z,t-3}$	–	0.024 (2.1)	0.028 (2.5)	–	0.077 (17.1)	0.035 (8.9)	−0.009 (−3.8)
$lcom_{zs,t-1}$	0.062 (8.2)	–	−0.339 (−42.9)	–	–	–	–
$lcom_{zs,t-2}$	−0.030 (−3.4)	0.017 (1.7)	0.170 (18.1)	–	–	–	–
$lcom_{zs,t-3}$	−0.039 (−5.2)	–	0.068 (8.7)	–	–	–	–
$c_{zs,t-1}$	–	0.080 (12.0)	0.069 (10.8)	−0.415 (−73.2)	–	0.005 (2.3)	–
$c_{zs,t-2}$	–	0.016 (2.0)	0.013 (1.8)	0.139 (21.0)	–	–	–
$c_{zs,t-3}$	−0.012 (−2.1)	–	–	0.112 (20.0)	–	−0.006 (−2.9)	–
R^2	0.089	0.098	0.100	0.148	0.107	0.114	0.100

Notes : **(a)** *Variables definition*: See table 2, note (b). **(b)** OLS estimates of first-differences of variables of interest on lagged variables of order 1, 2 and 3. We do not use higher-order lags in order to avoid superfluous moment conditions (Ziliak, 1997). All lagged variables are significant at least once in this table.

D Regressions by sector

Table D.1: Average plant size - Dynamic model by sector

Method	2SNLS		
$y^1_{zs,t}$	$P\Delta$		
Instruments	$\begin{pmatrix} y^1_{zs,t-3} \\ (\widetilde{\mathbf{x}}_{zs,t-j})_{j=0,1,2} \end{pmatrix}$		
	Manuf.	Trade	Services
y^2_{zst}	0.195	0.316	0.457
	(12.6)	(10.2)	(11.8)
lL_{zt}	0.333	0.257	0.121
	(12.9)	(8.0)	(3.9)
ls_{zt}	−0.231	−0.210	−0.250
	(6.4)	(−4.6)	(−4.1)
ld_{zt}	0.151	0.125	0.097
	(7.9)	(5.0)	(3.2)
$lcom_{zst}$	−0.593	−0.539	−0.613
	(−21.1)	(−11.5)	(−13.0)
c_{zst}	−0.107	−0.067	−0.001
	(−9.7)	(5.0)	(−0.1)
$y^1_{zs,t-1}$	0.786	0.754	0.748
	(16.2)	(8.5)	(9.3)
Sargan	0.064	0.824	0.057
$AC(1)$	−0.567	−0.559	−0.598
	($<10^{-4}$)	($<10^{-4}$)	($<10^{-4}$)
$AC(2)$	0.083	0.051	0.091
	($<10^{-4}$)	($<10^{-4}$)	($<10^{-4}$)
$AC(3)$	−0.067	0.029	−0.012
	(0.37)	(0.03)	(0.35)
Observations	39932	10270	11462

Notes : **(a)** *Variables definition*: See table 2, note (b). **(b)** $P\Delta$: See table 6, note (d). **(c)** *Significativity*: See table 6, note (c). **(d)** *Sargan*: See table 5, note (d).

Table D.2: Number of plants - Dynamic model by sector

Method	2SNLS		
$y^1_{zs,t}$	$P\Delta$		
Instruments	$\begin{pmatrix} y^1_{zs,t-3} \\ (\widetilde{\mathbf{x}}_{zs,t-j})_{j=0,1,2} \end{pmatrix}$		
	Manufacturing	Trade	Services
$y^1_{zs,t-1}$	-0.036 (-5.3)	-0.051 (-3.3)	-0.032 (2.1)
lL_{zt}	0.180 (3.7)	0.289 (4.0)	0.445 (5.2)
ls_{zt}	-0.076 (-3.1)	-0.067 (-1.3)	-0.191 (-2.5)
ld_{zt}	0.066 (2.8)	0.183 (3.4)	0.183 (3.1)
$lcom_{zst}$	0.952 (5.8)	0.678 (4.4)	0.330 (1.9)
$y^2_{zs,t-1}$	0.727 (16.1)	0.561 (9.4)	0.505 (6.6)
Sargan	.90	.56	.26
$AC(1)$.	-0.557 $(<10^{-4})$	-0.532 $(<10^{-4})$	-0.523 $(<10^{-4})$
$AC(2)$	0.047 $(<10^{-4})$	0.032 (0.02)	0.037 (0.004)
$AC(3)$	0.005 (0.58)	0.029 (0.06)	-0.013 (0.36)
Observations	27749	8706	9644

Notes : **(a)** *Variables definition*: See table 2, note (b). **(b)** $P\Delta$: See table 6, note (d). **(c)** *Significativity*: See table 6, note (c). **(d)** *Sargan*: See table 5, note (d). **(e)** Instruments do not comprise $lcomzs_{-1}$ and $lcomzs_{-2}$. Their validity is rejected.