

The Dynamics of Local Employment in France

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

First version: July 2002
This version: January 2004[§]

Abstract

We study the impact of local economic structure on employment dynamics. Local employment is decomposed into the product of the average plant size and the number of plants in the area and industry. We estimate the dynamics of both components simultaneously using French yearly data on 36 industries and 341 areas between 1984 and 1993. The careful specification of short-run dynamics and the control for fixed effects and endogeneity are shown to be critical in the empirical model. First, static externalities are prevalent compared to dynamic ones. Moreover, whereas area-and-industry effects explain most of the spatial variation of plant size, the local number of plants is mainly driven by the current local economic structure. Policies targeted towards plant creation should thus be more efficient. For instance, large areas endowed with a small number of even size industries have both more and larger plants. A large number of plants heterogenous in size also increases the average plant size, but the number of plants is higher when plants have similar size.

*CERAS-ENPC. CNRS researcher also affiliated with the CEPR. 48, boulevard Jourdan - 75014 Paris, France. E-mail: combes@enpc.fr. <http://www.enpc.fr/ceras/combes/>.

[†]INRA-Paris Jourdan and CREST-INSEE. Also affiliated with the CEPR and IZA. 48, boulevard Jourdan - 75014 Paris, France. E-mail: tmagnac@ens.fr. <http://www.inra.fr/ESR/UR/lea/magnac.htm>

[‡]Université de Paris 1 and CREST-INSEE. Also affiliated with the CEPR and IZA.

[§]Financial support from the Commissariat Général du Plan and NATO (Combes' advanced fellowship grant) are gratefully acknowledged. We thank Eve Caroli, Eric Maurin, the editor, two anonymous referees as well as seminar participants at Ammersee, Paris Jourdan, and PUC-Rio for their useful comments. The usual disclaimer applies.

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 fundamental question 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 [2]) 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 third level of the European Nomenclature of Territorial Units for Statistics (NUTS 2, 210 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. The former is 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 because of the large persistence of unemployment.

This study finds its roots in the seminal papers by Glaeser, Kallal, Scheinkman and Schleifer [11] and Henderson, Kuncoro and Turner [16]. 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.* [11] also consider the impact of local competition (“Porter externalities”). As surveyed by Combes and Overman [6], this has been replicated in various countries,¹ though 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 [14]. 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, these methods allow us to distinguish area-and-industry effects (capturing for instance the role of physical geography,

¹Among which Combes [3] in France using the same data though at the starting and final periods only.

climate, endowments, institutions, culture, etc) from local and time dependent externalities linked to the local economic structure. Furthermore, endogeneity can be properly controlled for. Second, it allows us to state whether externalities are dynamic and have a long lasting impact (“history matters”) or whether they are mainly static and influence local growth in the short run only. This old debate in urban economics is critical for assessing economic policies: Altering the local economic structure should be much more rapidly efficient in the second case but also less lasting.

Our approach relies on dynamic panel data analysis in order to distinguish between area-and-industry effects and economic structure and between static and dynamic externalities while controlling for endogeneity. As it is based on French data, it provides interesting comparisons between European and 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 the non-linear effect of industry employment that enters the specification both in logarithms and levels. To put it another way, if the model were specified only in logarithms, the specialization effect could not be identified from the initial industry employment effect (Combes, [3]). This makes interpretations problematic when both effects act in opposite directions as in Henderson *et al.* [16], and is the reason why we prefer not to use this specification. We also consider among explanatory variables two measures of diversity (on top of 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 in each area-and-industry cell. Instead of working with the dynamics of local industrial employment only, we simultaneously study the dynamics of 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 may simultaneously increase the size of existing plants and drive new firms into the market.

We use data on employment in 36 industries covering manufacturing, trade, and services, available for the 341 French continental “*zones d’emploi*” (“employment areas”). 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 auto-

correlation parameter is larger for plant size than for the number of plants. Moreover, whereas area-and-industry effects explained most of the spatial variation of plant size, the number of plants 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 also show that larger areas have more plants and larger ones. This correspond to agglomeration economies associated with the global size of the local economy (demand and cost linkages or technological spillovers that are not industry specific). They imply that faster growing areas experience larger area-and-industry growth in average plant size and plant creation. Interestingly, we refine the verdict regarding the impact of the industrial composition of the local economy on local employment. Whereas the number of locally active industries does not need to be large, industries of comparable size increase both the average plant size and the number of plants. An interpretation would be that technological spillovers can work across industries but do not extend to all industries. Similarly, intermediate inputs at the origin of demand and cost linkages would not be necessarily numerous in each industry but equally important. Hence, in both cases, the optimal structure would be small groups of even size industries.

Last, the impact of local competition is shown to be nonlinear. 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 favor the existence of 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 is larger in places where plants are of even size. All of these results clearly shed new light on the local economic structure that is the most favorable to local employment 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 of 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. We develop in this section a simple framework that helps identifying and clarifying the simplifying assumptions made in the literature. Next, we show how to extend it so as 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:

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

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:

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

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

Let us focus our attention on the impact on employment growth, $\frac{dL_z}{L_z}$ of a productivity shock, $\frac{dA_z}{A_z}$. 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, ε the supply elasticity of labor, which is assumed to be a finite positive constant. In contrast, the capital supply elasticity is supposed to be infinite. Those assumptions are consistent with previous ones regarding mobility. By definition, we have:

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

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

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

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 positive local productivity shock. First, it 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 a 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 labor expand. If the labor supply elasticity is infinite (local unemployment is involuntary for instance), the wage does not change and the effect on output and employment is maximal and equal to $\sigma - 1$.

Therefore, this simple framework leads to clarification of some important points. 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. In the previous literature, it is assumed that a positive productivity shock always induces employment expansion.

Glaeser *et al.* [11] 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:

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

that is to say, a mark-up over the marginal cost that depends on the number of firms. 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:

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

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 created and enter the market. In this case an additional 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, reflected 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, $\frac{d\bar{L}_z}{\bar{L}_z}$, where $\bar{L}_z = L_z/N_z$ 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:

$$\frac{dL_z}{L_z} = \frac{d\bar{L}_z}{\bar{L}_z} + \frac{dN_z}{N_z}, \quad (7)$$

Using equation (6) and (7), the impact of local externalities on firm size, $\frac{d\bar{L}_z}{\bar{L}_z}$, is given by:

$$\frac{d\bar{L}_z}{\bar{L}_z} = \varepsilon(\sigma - 1)\mu \left(\frac{dA_z}{A_z} + \frac{1}{\sigma N_z - 1} \frac{dN_z}{N_z} \right) - \frac{dN_z}{N_z}. \quad (8)$$

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). Also, 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 where profits are equal to zero.² By plugging the equilibrium price given by equation (5) into 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:

$$\frac{d\bar{L}_z}{\bar{L}_z} = (\varepsilon - 1)(\sigma - 1)\lambda \frac{dA_z}{A_z}, \quad (9)$$

²The assumption of fixed costs paid at each 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:

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

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

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

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). When σ is greater than 1, the demand increase because of decreasing prices is large enough to compensate the better efficiency of labor. It therefore induces increases both in total employment and in the number of firms. 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 (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 Vector AutoRegressive (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 (see section 4). 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. Notice that the

simultaneous estimation of the dynamics of firm size and the number of firms allows us to identify more effects than is usually achieved 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 employment 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 neighbouring regions. Indeed, as soon as trade between regions is properly modelled, the equilibrium price depends not only on local productivity but also on productivity of all trading partners. Derived specifications, as in, for instance, Combes and Lafourcade [4] for Cournot competition or Hanson [13] and Redding and Venables [18] for monopolistic competition, are much more intricate and beyond the scope of this literature. Again, the purpose of this paper is to provide stylized facts on local growth dynamics and not to estimate a particular model of trade and technology diffusion.

3 A Brief Look at the Data

We now describe the data we use and the variables that enter the VAR models we estimate. Data come from a survey about plant employment (*“Enquête Structure des Emplois”*) collected by the French National Institute of Statistics and Economic Studies (INSEE). This yearly survey reports the employment level of each plant between 1984 and 1993, the area where the plant is located and the industry it belongs to. We prefer to use data on plants rather than firms since the actual location of production is known. 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. We therefore limit our analysis to externalities resulting from the interaction of plants within and across local regional/industry clusters. One of the drawback of the survey is that it only includes all French-metropolitan plants that employ more than 20 employees. Comparable sampling procedures are used to study employment dynamics in U.S. cities. We leave to future research the assessment of whether our results are affected by this selection of plants.

We aggregate individual data on employment per plant by industry, area and date. We exclude agriculture and non-market services and group production activities into 36 different manufacturing, trade, and service industries. Public employment is excluded as its spatial distribution might respond to different incentives than private employment. The area code distinguishes 341 geographic units or employment areas (*“zones d’emploi”*). This INSEE spatial classification comprehensively covers the French territory, including both rural and urban areas. On average, areas spread over 1,570 km², which is fairly small (equivalent to splitting the continental U.S. territory

in more than 4,700 units). The definition of employment areas, performed by INSEE, is based on the observation of workers’ daily migrations. Given the number of areas, it aims at maximizing the number of within-area migrations between home and work. Hopefully, it 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 industry s at time t is characterized by the pair of variables (\bar{L}_{zst}, N_{zst}) , where $\bar{L}_{zst} = L_{zst}/N_{zst}$ is the average size of all plants located in area z and operating in industry s at time t , L_{zst} is total employment in cell (z, s, t) and N_{zst} the number of plants. We adopt a logarithmic specification and denote $\bar{\ell}_{zst} = \ln(L_{zst}/N_{zst})$ and $n_{zst} = \ln(N_{zst})$. It has the double advantage of making the distribution of these variables closer to a normal distribution and allowing for 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-industry (z, s) at time t) that have been “active” (i.e. $L_{zst} > 0$) since year $t - 1$, or since $t - 2$, and so on, down until $t = 1$, the first observation period (1984). Local areas and industries 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 industries 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 according to the off-diagonal elements of Table 1. Lastly, Combes [3] shows, yet in a single-equation model, that selectivity biases are limited in size.

3.2 Explanatory Variables

First, MAR externalities are reflected in our framework in the autoregressive dynamics of the series. Assume for instance that the process is AR(1). The auto-regressive parameter in the series of average plant size (resp. number of plants) indicates whether the growth of average plant size is larger if average plant size is larger in the area. Strictly speaking MAR externalities are observed only if the autoregressive coefficient is larger than 1. Yet, it would imply explosive employment

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

year (t)	total	$t-1$	$t-2$	$t-3$	$t-4$	$t-5$	$t-6$	$t-7$	$t-8$	$t-9$	$t-10$
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 follows: In 1984, the sample is composed by 7786 pairs (z, s) of local areas and industries; In 1985, 719 pairs (z, s) entered the sample and were not there in 1984 while 7603 pairs (z, s) are active and were in the sample in 1984, for a total of 8322, etc.

dynamics as some areas end up with no activity at all while others infinitely expand. If the autoregressive parameter is between 0 and 1, some inertia in the dynamics is observed as well as convergence to the long-run target. Average plant size and number of plants remain larger where they were larger, but spatial disparities in these variables decline (growth is larger where the variables take their lowest values), a phenomenon also called “mean-reversion”. Notice that some authors also interpret such non explosive dynamics when the autoregressive parameter is large as evidence of MAR externalities.

Besides the autoregressive dynamics reflecting MAR externalities, two groups of determinants of employment are comprised in the list of explanatory variables. The first group contains variables measuring urbanization externalities, namely:

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

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

This first variable, frequently used in the literature, captures global urbanization externalities, those related to the local market size and not to the industrial composition of the area.

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

$$s_{zt} = \ln(S_{zt}).$$

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

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

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

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

The second group of control variables measures Porter effects by characterizing the degree of competition between plants in the same industry. Contrary to urbanization externalities, which are indexed by area and period, but not by industry, local competition variables vary across areas, industries and dates. We consider the following two indicators of local competition:

1. The dispersion of local employment between plants within an industry as measured by the opposite of the logarithm of the Herfindahl index of within area-and-industry concentration:

$$comp_{zst} = -\ln \left[\sum_{i \in I_{zst}} \left(\frac{L_{it}}{L_{zst}} \right)^2 \right],$$

where L_{it} is the i^{th} plant size at period t and where we denote as I_{zst} the set of all plants i operating in area z and industry 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 industries (Encaoua and Jacquemin [10]).

2. An indicator of total absence of competition within an area and industry:

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

This choice of variables is justified by the survey of the literature presented in the introduction and the model detailed in section 2. Notice, however, that few studies simultaneously consider different measures of local diversity and competition. It is also 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 non-parametrically identified because of the colinearity between the dependent variables $\ln(L_{zst}/N_{zst})$ and $\ln(N_{zst})$ and the market size indicator $\ln L_{zt}$. An alternative sometimes adopted in the literature consists in introducing it in levels, but this makes its interpretation difficult as Combes [3] shows.

Table 2 provides descriptive statistics for all variables. Notice in particular that about 20% of all area, industry and period cases are mono-establishment ($mono_{zst} = 1$). Still, the local number of plants (in logs) is considerably more variable³ than average plant size. This is already an indication that a significant fraction of the dynamics of local employment should be explained by plant creation and destruction.

Table 2: Descriptive statistics

	Av.	Std.	Min	Max
$\bar{\ell}_{zst}$	4.18	0.76	2.99	10.12
n_{zst}	1.49	1.16	0	7.54
ℓ_{zt}	9.51	1.08	6.51	13.59
s_{zt}	3.22	0.24	1.79	3.58
div_{zt}	2.37	0.42	0.34	3.12
$comp_{zst}$	1.16	0.95	0	6.33
$mono_{zst}$	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:* $\bar{\ell}_{zst}$: $\ln(\text{plant average employment})$; n_{zst} : $\ln(\text{number of plants})$; ℓ_{zt} : $\ln(\text{total employment in the area})$; s_{zt} : $\ln(\text{number of industries with at least one plant with more than 20 employees in the area})$; div_{zt} : $\ln(\text{index of local diversity between industries})$; $comp_{zst}$: $\ln(\text{index of local competition within industry})$; $mono_{zst}$: dummy variable for there existing a monopoly in area z at time t

We conclude this section by noticing that time and industry dummies explain about one fourth of the total variance of dependent variables $\bar{\ell}_{zst}$ and 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 subtract from all variables their mean within period and industry cells.⁴

³As measured by the coefficient of variation: standard deviation divided by arithmetic mean.

⁴For example, let Z_{st} be the set of indices z for local areas where there exists a positive number of active firms in industry s at time t . We replace $\bar{\ell}_{zst}$ by

$$\bar{\ell}_{zst} - \frac{1}{\#Z_{st}} \sum_{z' \in Z_{st}} \bar{\ell}_{z'st}.$$

We proceed in the same way for variables n_{zst} , $comp_{zst}$ and $mono_{zst}$.

Variables ℓ_{zt} , s_{zt} and div_{zt} are simply detrended. For example, if Z is the set of all area indices, we replace ℓ_{zt} by

$$\ell_{zt} - \frac{1}{\#Z} \sum_{z' \in Z} \ell_{z't}.$$

Finally, area-and-industry effects are included in all specifications. This first controls for differences in the area-size of the geographical units we consider. A corollary is that variables as employment or number of plants can be interpreted in terms of spatial density. Second, these effects capture all area-and-industry time invariant effects that are possibly omitted. This is particularly important in this kind of approach for instance because areas are considered as closed-economies: Area- and-industry effects control at least for their relative location, the proximity to foreign countries, and thus more generally for physical geography. They can also be proxies for permanent spatial disparities in public endowments (such as airports), technology, or institutions, the effects of which are furthermore allowed to depend on the industry.⁵

4 Econometric Analysis

Let $\mathbf{y}_{zst} = (\bar{\ell}_{zst}, n_{zst})'$. We model the dynamics of vector \mathbf{y}_{zst} using a Panel Vector AutoRegressive (PVAR) specification of the form:

$$\mathbf{y}_{zst} = A(L)\mathbf{y}_{zs,t-1} + B(L)\mathbf{x}_{zst} + \mathbf{u}_{zs} + \mathbf{v}_{zst} \quad (12)$$

where $A(L)$ and $B(L)$ are matrix polynomials in the lag operator L and where \mathbf{v}_{zst} is a vector of random shocks. Dependent variables are correlated and the variance-covariance matrix of \mathbf{v}_{zst} is not supposed to be diagonal though it is supposed to be constant over time. Moreover, we allow for area-and-industry effects in both equations: $\mathbf{u}_{zs} = (u_{1zs}, u_{2zs})'$. If $A(L)$ and $B(L)$ are of degree 0, the model is static and both variables are explained by the current local economic structure only. It corresponds to equations (9) and (10) written for each industry s and date t . As detailed below, the more general dynamic model corresponding to system (12) with $A(L)$ and $B(L)$ of higher degree may be obtained by assuming an autoregressive structure of errors. A discussion of the univariate properties of the series, of details of the specification search for the best econometric model and of instrumental regressions, is available in Combes et al. [5]. It justifies in a much more precise way our preferred specification that is reported here, and the choice of the degree of the polynomials $A(L)$ and $B(L)$ for instance.

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

$$\bar{\ell}_{zst} = A_{11}(L)\bar{\ell}_{zs,t-1} + A_{12}(L)n_{zst} + B_1(L)\mathbf{x}_{zst} + u_{1zs} + \eta_{1zst} \quad (13)$$

$$n_{zst} = A_{21}(L)\bar{\ell}_{zs,t-1} + A_{22}(L)n_{zs,t-1} + B_2(L)\mathbf{x}_{zst} + u_{2zs} + \eta_{2zst} \quad (14)$$

where random shocks η_{1zst} and η_{2zst} are now uncorrelated and where $A_{ij}(L)$ and $B_i(L)$ are scalar polynomials in the lag operator. This recursive form is justified by the theoretical argument that

⁵Yet, these effects are fixed over time and do not capture time-varying effects of such local characteristics.

employment decisions are taken conditional on the number of active plants. It is indeed reasonable to assume that plant employment adjustments are far less costly than plant creations or destructions. Hence, if there exists an instantaneous causality between $\bar{\ell}_{zst}$ and n_{zst} , it is likely to be directed from the latter variable to the former one. This corresponds to equations (8) and (10) of the theoretical model (section 2) while system (12) would correspond to equations (9) and (10). Yet, this is a matter of interpretation only as the estimation of the two-equation system is consistent whatever the interpretation is, regarding the direction of instantaneous causality.

As shocks are now uncorrelated, we can estimate the two equations separately using the same methodology as for univariate dynamic models of panel data.

4.1 Average Plant Size ($\bar{\ell}_{zst}$)

We tried many different specifications of equations (13) and (14), for different numbers of lags and different assumptions on the dynamics of the error components η_{1zst} and η_{2zst} . The most parsimonious specification, and yet robust to specification testing, for plant size is displayed in Table 3. The estimated model explains $\bar{\ell}_{zst}$ by current values of the number of plants (n_{zst}) and the control variables (x_{zst}), and by these variables plus average plant size lagged only once:

$$\bar{\ell}_{zst} = \rho_1 \bar{\ell}_{zs,t-1} + \alpha_{10} n_{zst} + \alpha_{11} n_{zs,t-1} + \mathbf{x}'_{zst} \beta_{10} + \mathbf{x}'_{zs,t-1} \beta_{11} + u_{1zs} + \eta_{1zst}. \quad (15)$$

The first column of Table 3 reports the parameters of model (15) estimated using the technique of Anderson and Hsiao (see Arellano [1] or Hsiao [17]). The model is first-differenced to eliminate the area-and-industry (time-invariant) effect.⁶ The parameters are then estimated by instrumental variables (IV) using lagged values of the RHS variables (in levels) as instruments. The choice of instruments depends on the assumed order of autocorrelation of the error term in the differenced model ($\Delta\eta_{1zst}$) and on assumptions regarding to the exogeneity of these variables. We here assume that n_{zst} and $x_{zst} = (\ell_{zt}, s_{zt}, div_{zt}, comp_{zst}, mono_{zst})'$ are weakly exogenous with respect to η_{1zst} , i.e.

$$\mathbb{E}(\eta_{1zst} | n_{zst}, \mathbf{x}_{zst}, n_{zs,t-1}, \mathbf{x}_{zs,t-1}, \dots) = 0,$$

and that η_{1zst} is at most a Moving Average process of order 1 (MA(1)), so that $\Delta\eta_{1zst}$ is MA(2) and is thus orthogonal to the past of $\bar{\ell}_{zst}$ prior to $t - 2$. Hence:

$$\mathbb{E}(\Delta\eta_{1zst} | \bar{\ell}_{zs,t-3}, \bar{\ell}_{zs,t-4}, \dots, n_{zst}, \mathbf{x}_{zst}, n_{zs,t-1}, \mathbf{x}_{zs,t-1}, \dots) = 0.$$

In the estimation reported in the first column of Table 3 we used the following set of instruments:

$$\bar{\ell}_{zs,t-3}, n_{zst}, \mathbf{x}_{zst}, n_{zs,t-1}, \mathbf{x}_{zs,t-1}, n_{zs,t-2}, \mathbf{x}_{zs,t-2}.$$

⁶Time differentiation is more appropriate than within-estimation when the RHS variables are weakly endogenous (predetermined). It is the case here as the RHS of equation (15) contains lagged values of the dependent variable.

The Sargan test of overidentifying restrictions or, equivalently, the test of the validity of the instruments, comes out with a p-value well above 5%. This means that we do not reject the null hypothesis of exogeneity of the instruments at any reasonable level. Moreover, rows $AC(1)$, $AC(2)$ and $AC(3)$ report the estimates of the autocorrelation coefficients of order 1, 2 and 3 of the error term $\Delta\eta_{1zst}$. The significance probabilities (p-values of the t -test of nullity) indicate a strong rejection of the hypothesis that the first and second autocorrelations are zero but indicate that the autocorrelation of order 3 is not statistically different from zero, which is consistent with our initial assumption that $\Delta\eta_{1zst}$ is MA(2). To sum up, all specification tests show that our preferred specification of equation (13) fits data well and that our estimation method is likely to yield consistent parameter estimates.⁷

The estimate of the autoregressive coefficient ρ_1 at 0.878 is very precise and significantly different from both 0 and 1. Moreover, notice the alternating signs and the similar magnitude of the coefficients of control variables $n_{zs,t-1}, x_{zs,t-1}$ relative to $n_{zs,t-2}, x_{zs,t-2}$. This suggests estimating the following constrained form of model (15):

$$\bar{\ell}_{zst} - \rho_1 \bar{\ell}_{zs,t-1} = \alpha_1 (n_{zst} - \rho_1 n_{zs,t-1}) + (\mathbf{x}'_{zst} - \rho_1 \mathbf{x}'_{zs,t-1}) \beta_1 + u_{1zs} + \eta_{1zst}. \quad (16)$$

The estimates of the parameters of this equation are reported in the second column of Table 3. As this model is not linear, it is estimated by applying Two-Stage Non-Linear Least Squares (2SNLS) to the differenced model.⁸ We use the same set of instruments as before. Thanks to the constraints on the parameters, the estimates are very precise.⁹

Assuming a MA(1) structure for the transitory shock η_{1zst} :

$$\eta_{1zst} = \nu_{1zst} + \phi_1 \nu_{1zs,t-1},$$

parameter ϕ_1 can be recovered by using the residuals of the estimation reported in Table 3. Moreover, an estimate of the variance of ν_{1zst} can also be recovered from the estimated variance of these residuals. The moving average parameter ϕ_1 is very precisely estimated at a small value: -0.140 (t -stat = -12.8). This negative sign is consistent with the existence of measurement errors, which are plausible given the data sampling rule of recording the employment of the sole plants employing more than 20 workers. However, as the autocorrelation coefficient ϕ_1 is small enough and stands for

⁷Again, this is the result of the search for the best specification, as detailed in Combes et al. (2003).

⁸We exploit the bilinearity of the model as follows. Let z_{zst} be a vector of instruments. Parameters ρ_1 and $b_1 = (\alpha_1, \beta_1)$ can then be estimated by iterating until convergence the following algorithm: Given a pair $(\rho^{(n)}, b^{(n)})$ of parameter values, a new estimate $(\rho^{(n+1)}, b^{(n+1)})$ is computed by applying 2SLS to the regression of $\Delta\bar{\ell}_{zst}$ on variables $\Delta\bar{\ell}_{zs,t-1}$ and $\Delta x_{zst} - \rho^{(n)} \Delta x_{zs,t-1}$ instrumented by z_{zst} .

⁹Note that the reported standard errors are biased as they have been computed under the assumption that the estimate of ρ_1 is independent of the estimate of (α_1, β_1) and assuming away any serial and spatial dependence. However, we tried to implement some corrections to get a sense of their magnitude. These experiments indicate that one should not expect t -statistics to decrease by much more than a factor of 30%, which make the variables still largely significant. These results are available upon request.

Table 3: Average plant size - Dynamic model

	(1)	(2)
n_{zst}	0.281 (4.9)	0.287 (20.1)
ℓ_{zt}	0.257 (1.6)	0.268 (15.3)
s_{zt}	-0.093 (0.5)	-0.231 (8.7)
div_{zt}	0.166 (2.4)	0.136 (9.8)
$comp_{zst}$	-0.497 (-11.9)	-0.576 (26.7)
$mono_{zst}$	-0.002 (-0.1)	-0.063 (8.5)
$\bar{\ell}_{zs,t-1}$	0.878 (16.3)	0.809 (20.1)
$n_{zs,t-1}$	-0.257 (11.5)	-
$\ell_{z,t-1}$	-0.237 (-4.2)	-
$s_{z,t-1}$	0.239 (4.1)	-
$div_{z,t-1}$	-0.109 (-4.4)	-
$comp_{z,t-1}$	0.529 (16.3)	-
$mono_{z,t-1}$	0.072 (4.9)	-
Sargan test (p-value)	0.839	0.521
$AC(1)$	-0.579 ($<10^{-5}$)	-0.576 ($<10^{-5}$)
$AC(2)$	0.084 ($<10^{-5}$)	0.082 ($<10^{-5}$)
$AC(3)$	-0.003 (0.57)	0.004 (0.53)
Sample size	54664	54664

Notes: 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.

measurement errors, the selection of rather large establishments above 20 employees is not likely to bias the estimation too much.

It is finally possible to rewrite model (16) in the following way. Since the estimate of the autoregressive parameter ρ_1 is significantly different from one, one can divide both members of equation (16) by the lag polynomial $1 - \rho_1 L$ and obtain the following representation:

$$\bar{\ell}_{zst} = \alpha_1 n_{zst} + \mathbf{x}'_{zst} \beta_1 + \frac{u_{1zs}}{1 - \rho_1} + \frac{\eta_{1zst}}{1 - \rho_1 L}. \quad (17)$$

The dynamics of local average plant size $\bar{\ell}_{zst}$ is thus very simple: $\bar{\ell}_{zst}$ fluctuates in a stationary way around a moving target, that is:

$$\bar{\ell}_{zst} = \bar{\ell}_{zst}^{LT} + \varepsilon_{1zst}, \quad (18)$$

with

$$\bar{\ell}_{zst}^{LT} = \alpha_1 n_{zst} + \mathbf{x}'_{zst} \beta_1 + \frac{u_{1zs}}{1 - \rho_1} \quad (19)$$

and $\varepsilon_{1zst} = \frac{\eta_{1zst}}{1 - \rho_1 L}$ is an ARMA(1,1) stationary process.¹⁰

Equation (17) is an important result in itself as it shows the validity of the static Cournot model developed in the first section. The relationship between, on the one hand, average plant size $\bar{\ell}_{zst}$ and the number of plants n_{zst} and, on the other hand, the control variables \mathbf{x}_{zst} which are supposed to determine productivity shocks is exactly of the form of the theoretical equation (8). Innovation errors ν_{1zst} , or shocks driven by unobserved variables, shift $\bar{\ell}_{zst}$ away from its target $\bar{\ell}_{zst}^{LT}$. The return to the equilibrium is smooth but relatively slow, with an auto-regressive coefficient at $\rho_1 = 0.809$: After 10 years, an initial positive innovation shock still contributes to set $\bar{\ell}_{zst}$ above $\bar{\ell}_{zst}^{LT}$ by a fraction $(\rho_1)^{10} = 12\%$ of the initial shock. Impulse-response functions are commented below after the completion of the analysis for both series.

We report in Table 4 the variance-covariance decomposition induced by the estimation of equation (17). The long-term target $\bar{\ell}_{zst}^{LT}$ explains 85% of the total variance. This is by all means a large contribution, and it means that shocks ε_{1zst} are never big enough to move $\bar{\ell}_{zst}$ very far away from its equilibrium path $\bar{\ell}_{zst}^{LT}$. The contribution of the area-and-industry effects u_{1zs} to the variance of $\bar{\ell}_{zst}^{LT}$ is also found to be large (87%). Control variables n_{zst} , ℓ_{zt} , div_{zt} , $comp_{zst}$ and $mono_{zst}$ are thus found to determine the long-run path of $\bar{\ell}_{zst}$ only moderately (16%). The shares of variance explained do not sum to 1 because they are not orthogonal (the correlation is equal to -0.042).

¹⁰The variance of which is

$$\mathbb{V}(\varepsilon_{1zst}) = \mathbb{V}\left(\frac{\eta_{1zst}}{1 - \rho_1 L}\right) = \mathbb{V}\left(\frac{\nu_{1zst} + \phi_1 \nu_{1zs,t-1}}{1 - \rho_1 L}\right) = \frac{1 + 2\rho_1 \phi_1 + \phi_1^2}{1 - \rho_1^2} \mathbb{V}(\nu_{1zst}).$$

Table 4: Average plant size - Variance decomposition

$\frac{\mathbb{V}\bar{\ell}_{zst}^{LT}}{\mathbb{V}\bar{\ell}_{zst}}$	$\frac{\mathbb{V}(\alpha_1 n_{zst} + \mathbf{x}'_{zst}\beta_1)}{\mathbb{V}\bar{\ell}_{zst}^{LT}}$	$\frac{\mathbb{V}\left(\frac{u_{1zs}}{1-\rho_1}\right)}{\mathbb{V}\bar{\ell}_{zst}^{LT}}$	$\text{Corr}\left(\alpha_1 n_{zst} + \mathbf{x}'_{zst}\beta_1, \frac{u_{1zs}}{1-\rho_1^2}\right)$
0.852	0.161	0.871	-0.042

4.2 Number of Plants (n_{zst})

In this section, we replicate the previous methodology to analyze the dynamics of the number of plants by area-and-industry. We start with the following specification:

$$n_{zst} = \rho_2 n_{zs,t-1} + \alpha_{21} \bar{\ell}_{zs,t-1} + \alpha_{22} \bar{\ell}_{zs,t-2} + \mathbf{x}'_{zst} \beta_{20} + \mathbf{x}'_{zs,t-1} \beta_{21} + u_{2zs} + \eta_{2zst}, \quad (20)$$

where the coefficient of the variable $mono_{zst}$ in β_{20} is constrained to be zero to avoid explaining n_{zst} by a function of itself. We assume that

$$\mathbb{E}(\Delta \eta_{2zst} | \bar{\ell}_{zs,t-3}, \bar{\ell}_{zs,t-4}, \dots, n_{zst}, \mathbf{x}_{zst}, n_{zs,t-1}, \mathbf{x}_{zs,t-1}, \dots) = 0$$

and use IV on the differenced model with instruments

$$\bar{\ell}_{zs,t-3}, n_{zst}, \mathbf{x}_{zst}, n_{zs,t-1}, \mathbf{x}_{zs,t-1}, n_{zs,t-2}, \mathbf{x}_{zs,t-2}.$$

The first column of Table 5 reports these estimates. We again observe that the coefficients of each variable and its lag have opposite signs, and therefore reestimate using the same technique as before, the following constrained specification:

$$n_{zst} - \rho_2 n_{zs,t-1} = \alpha_2 (\bar{\ell}_{zs,t-1} - \rho_2 \bar{\ell}_{zs,t-2}) + (\mathbf{x}'_{zst} - \rho_2 \mathbf{x}'_{zs,t-1}) \beta_2 + u_{2zs} + \eta_{2zst}. \quad (21)$$

Sargan statistics indicate that overidentifying restrictions cannot be rejected at any 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 levels.

We can thus safely conclude that the dynamics of plant creations and destructions is also correctly captured by a model assuming random oscillations around a time-dependent long-run target:

$$n_{zst}^{LT} = \alpha_2 \bar{\ell}_{zs,t-1} + \mathbf{x}'_{zst} \beta_2 + \frac{u_{2zs}}{1-\rho_2} \quad (22)$$

and the difference between the actual number of plants and its long-term target, $\varepsilon_{2zst} = n_{zst} - n_{zst}^{LT}$, is again an ARMA(1,1) process:

$$\varepsilon_{2zst} = \frac{\eta_{2zst}}{1-\rho_2 L}.$$

Table 5: Number of plants - Dynamic model

	(1)	(2)
$\ell_{zs,t-1}$	-0.041 (-1.7)	-0.042 (-10.1)
ℓ_{zt}	0.191 (1.9)	0.243 (18.5)
s_{zt}	-0.107 (-0.8)	-0.108 (-5.3)
div_{zt}	0.082 (1.5)	0.096 (9.3)
$comp_{zst}$	0.818 (32.1)	0.831 (37.1)
$n_{zs,t-1}$	0.829 (15.6)	0.710 (27.4)
$\bar{\ell}_{zs,t-2}$	0.036 (5.1)	-
$\ell_{z,t-1}$	-0.219 (-6.4)	-
$s_{z,t-1}$	0.092 (2.4)	-
$ld_{z,t-1}$	-0.086 (-4.5)	-
$comp_{z,t-1}$	-0.692 (-16.1)	-
Sargan	0.300	0.122
$AC(1)$	-0.571 ($<10^{-5}$)	-0.561 ($<10^{-5}$)
$AC(2)$	0.062 ($<10^{-5}$)	0.058 ($<10^{-5}$)
$AC(3)$	-0.001 (0.91)	-0.003 (0.71)
Obs	54664	54664

The autoregressive coefficient ρ_2 is estimated to be 0.71 which implies a relatively high level of persistence, lower than the value that was found for average plant size, however. The estimate of the MA parameter ϕ_2 is -0.123 (with a t -statistic of -11.3). It is significant but small and negative, which is again consistent with the existence of rather moderate measurement errors.

We report in table 6 the appropriate variance-covariance decomposition. The variance of the long-term target n_{zst}^{LT} explains about 74% of the total variance, which again means that the series n_{zst} never departs very far from its long-term target. Now, contrary to what was obtained for the average plant size, most of the long-term variance ($\mathbb{V}n_{zst}^{LT}$) is explained by past average plant size $\bar{\ell}_{zs,t-1}$ and contemporaneous control variables \mathbf{x}_{zst} (over 96%). We thus find no significant contribution of an area-and-industry effect, which are, however, correlated to the other variable effects (correlation at 0.263) more strongly than it was for average plant size.

Table 6: Number of plants - Variance decomposition

$\frac{\mathbb{V}n_{zst}^{LT}}{\mathbb{V}n_{zst}}$	$\frac{\mathbb{V}(\alpha_2\bar{\ell}_{zs,t-1} + \mathbf{x}'_{zst}\beta_2)}{\mathbb{V}n_{zst}^{LT}}$	$\frac{\mathbb{V}\left(\frac{u_{2zs}}{1-\rho_2}\right)}{\mathbb{V}n_{zst}^{LT}}$	$\text{Corr}\left(\alpha_2\bar{\ell}_{zs,t-1} + \mathbf{x}'_{zst}\beta_2, \frac{u_{2zs}}{1-\rho_2}\right)$
0.744	0.965	0.004	0.263

Finally, we repeated the analysis for manufacturing, trade and services separately. The results are reported in Tables D.1 and D.2 in Appendix B. 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.

4.3 Interpretations

In this section we put into perspective the previous estimation results. 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 in order to obtain a complete view of the local employment and firm creation dynamics (Davis, Haltiwanger and Schuh, [7]).

We start with two important general conclusions that stem from our preferred empirical models which describe both dependent variables as stationary noises around a long run target. First, the variance of such a noise accounts for about 15% of the total variance of dynamics of average plant size. Regarding the number of plants, this ratio is slightly larger since the noise accounts for about 25% of the total variance. Temporary shocks thus appear to be relatively secondary compared to the effects of current local conditions. In other words, our data support economic geography models based on static trade models under imperfect competition and where the role of static

externalities in employment or production growth is emphasized. Our results therefore differ from those obtained by Henderson [14] for the U.S., who finds much more important fluctuations.

Second, we show that regarding plant size, area-and-industry effects significantly and strongly matter in the determination of the long-run target. Therefore, plant size seems to be weakly affected by local externalities occurring because of interactions between agents. By contrast, local fixed endowments such as those of physical geography (proximity to a border, to an ocean, presence of mountains, of mines etc), local public goods, institutions, or permanent comparative advantages seem to play an important role and is possibly industry dependent. 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 though these externalities are shown to be also moderately correlated with area-and-industry effects.

Dynamic effects are not absent, however. The persistence of shocks is measured by the autoregressive coefficient, ρ . It gives evidence of a mean reversion process. When estimated separately for manufacturing, trade, and services (Tables D.1 and D.2 in the Appendix), it is around 0.7-0.8 for plant size and 0.5-0.7 for the number of plants. For average plant size, we find little difference between sectors. By contrast, the autoregressive coefficient of the number-of-plants series is more variable across sectors. Trade and services differ from manufacturing by a weaker persistence in their creation/destruction of plants. At face value, Marshall-Arrow-Romer externalities are not observed, which would be the case if the autoregressive coefficient was larger than one, though large 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 presents larger plant size and number of plants, this remains the case for a number of periods. Yet, convergence across areas occurs in the long-run, conditionally on the local economic structure at that time.

Impulse-response functions are graphed in Figure 1. In the top panel, we generated an innovation shock affecting average plant size at period 0 only and which magnitude is equal to the estimate of one standard error of the innovation shock in this series. In the bottom panel, we generated an innovation shock affecting the number of plants at period 0 only. We then computed the future trajectories of both series in both cases by assuming that the local economic structure remains constant over time. These diagrams are thus describing fluctuations around long-run targets. Years are along the x -axis, while the y -axis should be read as the relative magnitude of the impacts of the shocks on the logarithms of dependent variables (about 17% in the top panel for instance at period 0 for the average plant size) and the number of plants. Both diagrams exhibit quite strong responses at period 0 and convergence only occurs at a rather slow rate. After 10 years, the remaining effect of the innovation on average plant size is equal to 2% while it is equal to 0.5% for the number of plants. It is also noticeable that shocks on the number of plants affect average plant size much more than the reverse. Moreover, since results can be read in first differences, comparable

statements can be advanced for growth rates. Large external or internal growth rates at a given period imply large growth rates in the future, even more for plant size and for manufacturing, at least in the short run. Yet, convergence in growth rates occurs in the long run.

Concluding the analysis of dynamic effects, the larger persistence of plant size and the larger share of variance of the series of plant size which is explained by the area-and-industry effect have an important policy implication. Economic policies affecting the local economic structure would be more efficient in creating new plants, that is external growth, than those policies trying to influence the growth of existing plants. Moreover, they would also be more effective in trade and services than in manufacturing.

Contrary to previous studies, we are able to distinguish these short-run dynamic effects from the impact on the long-run target of the local economic structure conditional on the presence of area-and-industry effects. We analyze the respective impact of each variable to which economic policies could be targeted and from now on growth is interpreted as growth of the long-run target. Let us start by the impact on local growth of the global size of the local area (ℓ_{zt}). We find that growing areas experience both internal and external growth and thus the data exhibit global agglomeration economies. Faster growing areas where both technological spillovers might be stronger and where final and intermediate good markets are growing faster have larger growth in average plant size as well as stronger plant creation. This last point is in particular consistent with the idea of “nursery” cities developed by Duranton and Puga [9]. These 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 though activity moves towards other 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 also been used to describe the impact of local industrial diversity (div_{zt}) on local growth. To describe the local industrial composition, we also included the number of active industries (s_{zt}) as an additional explanatory variable next to the diversity index. Both variables are always significant in all industries. Diversity has a positive effect on plant size and on the number of plants. In contrast, the number of operating industries in the area (s_{zt}) has a negative effect in both cases.

As a consequence, the message regarding the role of the industrial composition is refined. The most favorable local industrial structure would consist in a small number of industries but of roughly the same size, which would maximize both the area-and-industry number of plants and average plant size. In terms of agglomeration forces, this finding 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-industrial but would not extend to all industries. They would be maximized inside fairly small but balanced sub-groups of similar sized industries.

Ceteris paribus, the effect of the number of plants (n_{zst}) on plant size ($\bar{\ell}_{zst}$) is positive. In other words, employment in an area and industry grows more quickly than the number of plants. This can be interpreted in terms of pure local externalities (a larger number of plants favors innovation and then plant size) and of market based forces. In the latter case, using the imperfect competition set up developed in section 2 allows 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 prices which next increase demands. Conversely, the effect of the lagged plant size ($\bar{\ell}_{zst}$) on the number of plants (n_{zst}) is small but negative. This result 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 industry ($comp_{zst}$) has a strong and negative effect on plant size and has a strong and positive effect on the number of plants. The more concentrated across plants employment is, the larger average plant size is though the smaller the number of plants 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, $comp_{zst}$, 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 ($mono_{zst}$) decreases as regards plant size.¹¹

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. [11], competition is proxied by the (inverse) size of plant and results can be hardly compared to our approach. In Combes [3] only one competition variable, $comp_{zst}$, is considered. It has a negative effect almost everywhere except in a few service industries. Again, the inclusion of different competition variables here, and the fact that they are all significant (except the monopoly situation in services), allows for a more precise verdict. Competition seems to have non-linear effects on local employment in industries. Average plant size is maximized when the number of local plants is large, when they are definitely not in a situation of local monopoly (both being consistent with the idea of more competition) and when, simultaneously, the size of plants is uneven (which is associated with a lesser degree of competition). One can think about a situation in which the presence of some large leader firms increase the number of smaller surrounding plants. The size of the leaders would induce benefits due to economies of scale while the followers could for instance gain from cost and demand linkages. This view is also consistent with the idea of large plants doing research and

¹¹As already said, this variable has obviously been omitted from the equation explaining the number of plants.

development that benefits all other plants, a structure that would maximize technological spillovers and knowledge diffusion. The counterpart would be that the existence of plants of even size on the contrary is more favorable to the entry of new plants. This complex impact of local competition on employment would certainly deserve more investigation in the future, both theoretically and empirically.

5 Conclusion

We analyzed yearly data extracted from a survey about plant employment and we studied the impact of the local economic structure on local growth of plant size and number of plants in France between 1984 and 1993. The short-run dynamics is carefully modelled by controlling for area-and-industry effects and endogeneity and both 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 presented and more generally of urban economics and economic geography. Economic policy implications 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. However, 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 of 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 [8], for instance, propose a methodology that could be mixed with ours to evaluate the distance at which agglomeration forces operate. We did not analyze either the influence of local infrastructure and more generally of public goods available at the local level separately, which are simply controlled by area-and-industry effects as other endowments are. Finally, Henderson [15] using plant data in machinery and high-tech industries assesses the respective role of localization and urbanization economies on plant productivity, also controlling for fixed-effects and endogeneity. Merging both kinds of approaches could give a more complete view of the joint dynamics of production, productivity and employment. Research on these issues is certainly high on the agenda.

References

- [1] M., Arellano, Panel Data Econometrics, Oxford U.P.: Oxford, 2003.

- [2] R. Barro, X. Sala-I-Martin, *Economic Growth*, MIT Press, Cambridge, 1995.
- [3] P.-P. Combes, *Economic Structure and Local Growth: France, 1984-1993*, *Journal of Urban Economics* 47 (2000) 329-355.
- [4] P.-P. Combes, M. Lafourcade, *Transport Costs Decline and Regional Inequalities: Evidence from France*, CEPR Discussion paper, 2894 (2001).
- [5] P.-P. Combes, T. Magnac, J.-M. Robin, *The Dynamics of Local Employment in France*, CEPR discussion paper 3912 (2003).
- [6] P.-P. Combes, H.G. Overman, *The Spatial Distribution of Economic Activities in the European Union*, in: V. Henderson, J.-F. Thisse (Eds.), *Handbook of Urban and Regional Economics*, Vol. 4, North Holland, Amsterdam, 2004.
- [7] S. Davis, J. Haltiwanger, S. Schuh, *Small Business and Job Creation: Dissecting the Myth and Reassessing the Facts*, *Small Business Economics* 8 (1996) 297-315.
- [8] K. Desmet, M. Fafchamps, *The Changing Spatial Distribution of Economic Activity across U.S. Counties*, Oxford University mimeo (2001).
- [9] G. Duranton, D. Puga, *Nursery Cities: Urban diversity, process innovation, and the life cycle of products*, *American Economic Review* 91 (2001) 1454-1477.
- [10] D. Encaoua, A. Jacquemin, *Degree of Monopoly, Indices of Concentration and Threat of Entry*, *International Economic Review*, 21 (1980) 87-105.
- [11] E. L. Glaeser, H. Kallal, J. Sheinkman, A. Schleifer, *Growth in Cities*, *Journal of Political Economy*, 100 (1992) 1126-1152.
- [12] C. Gouriéroux, A. Monfort, *Statistics and Econometric Models*, Cambridge University Press, Cambridge 1999.
- [13] Hanson, G., *Market Potential, Increasing Returns, and Geographic Concentration*, revised version of the NBER Working Paper 6429 (2000).
- [14] J. V. Henderson, *Externalities and industrial development*, *Journal of Urban Economics*, 42 (1997) 449-470.
- [15] J. V. Henderson, *Marshall's Scale Economies*, *Journal of Urban Economics*, 53 (2003) 1-28.
- [16] J. V. Henderson, A. Kuncoro, M. Turner, *Industrial Development in Cities*, *Journal of Political Economy*, 103 (1995) 1067-1090.

[17] C. Hsiao, Analysis of Panel Data, Cambridge University Press, Cambridge 2003.

[18] S. Redding, A. Venables, Economic Geography and International Inequality, Journal of International Economics, 62 (2004) 53-82.

Appendix

A Theoretical Predictions

A.1 Perfect Competition

Differentiating the production function (1) yields:

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

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

$$\frac{dp_z}{p_z} = \alpha \frac{dw_z}{w_z} - \frac{dA_z}{A_z}. \quad (\text{A.2})$$

Equalization of labor productivity to real wage is given by:

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

and differentiation yields:

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

We then have five linear equations (3×2 , A.1, A.2, A.4), in five unknowns: $\frac{dY_z}{Y_z}$, $\frac{dL_z}{L_z}$, $\frac{dK_z}{K_z}$, $\frac{dp_z}{p_z}$, $\frac{dw_z}{w_z}$ 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:

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

where

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

is the marginal cost, f is the fixed 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

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

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:

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

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:

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

Using

$$\frac{dL_z}{L_z} = \frac{d\ell_z}{\ell_z} + \frac{dN_z}{N_z},$$

we obtain:

$$\frac{d\ell_z}{\ell_z} = \frac{\varepsilon(\sigma - 1)}{1 - \alpha + \varepsilon + \alpha\sigma} \left(\frac{dA_z}{A_z} + \frac{1}{\sigma N_z - 1} \frac{dN_z}{N_z} \right) - \frac{dN_z}{N_z}. \quad (\text{A.9})$$

A.3 Cournot Competition: Long Run

The first order condition (A.6) and free entry yield:

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

By differentiation,

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

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

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

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

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. $\frac{d\ell_z}{\ell_z}$ also needs $\varepsilon > 1$ to be positive

B Regressions by sector

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

	Manuf.	Trade	Services
n_{zst}	0.195 (12.6)	0.316 (10.2)	0.457 (11.8)
ℓ_{zt}	0.333 (12.9)	0.257 (8.0)	0.121 (3.9)
s_{zt}	-0.231 (6.4)	-0.210 (-4.6)	-0.250 (-4.1)
div_{zt}	0.151 (7.9)	0.125 (5.0)	0.097 (3.2)
$comp_{zst}$	-0.593 (-21.1)	-0.539 (-11.5)	-0.613 (-13.0)
$mono_{zst}$	-0.107 (-9.7)	-0.067 (5.0)	-0.001 (-0.1)
$y_{zs,t-1}^1$	0.786 (16.2)	0.754 (8.5)	0.748 (9.3)
Sargan	0.064	0.824	0.057
AC(1)	-0.567 ($<10^{-4}$)	-0.559 ($<10^{-4}$)	-0.598 ($<10^{-4}$)
AC(2)	0.083 ($<10^{-4}$)	0.051 ($<10^{-4}$)	0.091 ($<10^{-4}$)
AC(3)	-0.067 (0.37)	0.029 (0.03)	-0.012 (0.35)
Observations	39932	10270	11462

Notes : Equation (16) is estimated using non linear IV on the differenced model using instruments $\bar{\ell}_{zs,t-3}$, n_{zst} , \mathbf{x}_{zst} , $n_{zs,t-1}$, $\mathbf{x}_{zs,t-1}$, $n_{zs,t-2}$, $\mathbf{x}_{zs,t-2}$.

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

	Manufacturing	Trade	Services
$\bar{l}_{zs,t-1}$	-0.036 (-5.3)	-0.051 (-3.3)	-0.032 (2.1)
l_{zt}	0.180 (3.7)	0.289 (4.0)	0.445 (5.2)
s_{zt}	-0.076 (-3.1)	-0.067 (-1.3)	-0.191 (-2.5)
div_{zt}	0.066 (2.8)	0.183 (3.4)	0.183 (3.1)
$comp_{zst}$	0.952 (5.8)	0.678 (4.4)	0.330 (1.9)
$n_{zs,t-1}$	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) Equation (21) is estimated using non linear IV on the differenced model using instruments $\bar{l}_{zs,t-3}$, n_{zst} , \mathbf{x}_{zst} , $n_{zs,t-1}$, $\mathbf{x}_{zs,t-1}$, $n_{zs,t-2}$, $\mathbf{x}_{zs,t-2}$. (b) Instruments do not comprise $comp_{zs,t-1}$ and $comp_{zs,t-2}$. The test of their validity (exogeneity) was rejected.

