



HAL
open science

Spatial Concentration and Firm-Level Productivity in France

Philippe Martin, Thierry Mayer, Florian Mayneris

► **To cite this version:**

Philippe Martin, Thierry Mayer, Florian Mayneris. Spatial Concentration and Firm-Level Productivity in France. 2008. hal-01066174

HAL Id: hal-01066174

<https://sciencespo.hal.science/hal-01066174>

Preprint submitted on 19 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DISCUSSION PAPER SERIES

No. 6858

SPATIAL CONCENTRATION AND FIRM-LEVEL PRODUCTIVITY IN FRANCE

Philippe Martin, Thierry Mayer
and Florian Mayneris

*INTERNATIONAL TRADE AND
REGIONAL ECONOMICS*



Centre for **E**conomic **P**olicy **R**esearch

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP6858.asp

SPATIAL CONCENTRATION AND FIRM-LEVEL PRODUCTIVITY IN FRANCE

Philippe Martin, Paris School of Economics, University of Paris I and CEPR
Thierry Mayer, Paris School of Economics, University of Paris I, CEPPII and CEPR
Florian Mayneris, Paris School of Economics and Ecole des Hautes Etudes en
Sciences Sociales

Discussion Paper No. 6858
June 2008

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL TRADE AND REGIONAL ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Philippe Martin, Thierry Mayer and Florian Mayneris

June 2008

ABSTRACT

Spatial Concentration and Firm-Level Productivity in France*

This paper analyzes empirically the effect of spatial agglomeration of activities on the productivity of firms using French individual firm data from 1996 to 2004. This allows us to control for endogeneity biases that the estimation of agglomeration economies typically encounters. French firms benefit from localization economies, but not from urbanization economies nor from competition effects. The benefits generated by increased sectoral clustering, though positive and highly significant are modest and geographically very limited. The gains from clusters are also quite well internalized by firms in their location choice: we find very little difference between the geography that would maximize productivity gains and the geography actually observed.

JEL Classification: C23, R10, R11 and R12

Keywords: clusters, localization economies, productivity and spatial concentration

Philippe Martin
University of Paris 1 Panthéon-
Sorbonne
106-112 Bd de l'hôpital
75647 Paris CEDEX 13
FRANCE
Email: philippe.martin@univ-paris1.fr

Thierry Mayer
University of Paris 1 Panthéon-
Sorbonne
106-112 Bd de l'hôpital
75647 Paris CEDEX 13
FRANCE
Email: tmayer@univ-paris1.fr

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=116056

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=152707

Florian Mayneris
Paris School of Economics
48 bd Jourdan
75014 PARIS
FRANCE
Email: mayneris@pse.ens.fr

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=168305

* We thank Anthony Briant, Pierre-Philippe Combes and Gilles Duranton for helpful comments, and CEPREMAP for the financial assistance.

Submitted 29 May 2008

1 Introduction

In the late XIXth century, Alfred Marshall (1890) was fascinated by the concentration of cutlery and hosier industries in the regions of Sheffield and Northampton. If no natural advantages could explain the agglomeration of those industries in such places, which mechanisms could drive the formation of productive clusters? According to Marshall, industrial districts are in reality places where “the mysteries of the trade become no mysteries; but are as it were in the air and children learn many of them unconsciously...”. Beyond that “mystical approach”, three main sources of agglomeration externalities have been identified first by Marshall, and rediscovered later by Arrow and Romer¹:

- *Externalities on inputs market*: the concentration of producers of a given industry on a given territory generates an incentive for input suppliers to locate in the neighborhood. As a consequence, producers can share specialized services, save expensive transportation costs or manage more efficiently their purchases of inputs.
- *Externalities on the labor market*: industrial clusters favor the creation of pools of specialized workers, who acquire cluster-specific skills valuable to the firms.
- *Knowledge externalities*: industrial clusters facilitate the exchange of information and knowledge and seem to be a form of organization particularly favorable to technological and knowledge spillovers.

These industry-specific externalities are usually grouped under the concept of *localization economies*, as opposed to Jane Jacobs’ *urbanization economies*, which refer to the cross fertilizations of *different industries* on a given territory. In the rest of the paper, we will systematically distinguish those two types of agglomeration externalities.

The analysis of agglomeration economies is not only important to better understand the economic mechanisms at work at the local level, it has also potentially important policy implications. Since the end of the 1980’s, agglomeration economies have been used to justify cluster policies by national and local governments in Germany, Brazil, Japan, Southern Korea, Spanish Basque country or more recently France. Some of those policies are very costly. For example, 1.5 billions euros have been devoted to the “Competitiveness clusters” by the French government. Two separate questions deserve attention to clarify the policy debate. First, how large are the gains from agglomeration? In particular, how much does the productivity of a firm increase when other firms from the same sector or from another sector decide to locate nearby? Second, how much do firms internalize these gains when deciding where to locate? The answer to the first question should help understand how much economic gains can be expected from clusters. The answer to the second question should help understand whether there is a strong case for public intervention in favor of industrial clusters².

A disturbing feature of the existing literature is that the existence of agglomeration gains is such that one would be tempted to conclude that more agglomeration is always better for the productivity of firms. This does not look very plausible as congestion costs must necessarily appear and dominate at a certain level of agglomeration. If this was not so, one should also

¹So that the term Marshall-Arrow-Romer (MAR) externalities is often used.

²See Duranton, Martin, Mayer, and Mayneris (2008) for more detail about this.

conclude that the observed geography (where all firms of the same sector are not located in the same region) is vastly suboptimal. In this paper, based on French firm level data, we find that the gains from clustering do exist: a 10% increase of employment in neighboring firms of the same industry increases a firm productivity by around 0.4-0.5%. When we use a non linear specification, we find that the relation between productivity gains and agglomeration is bell shaped and we are able to estimate the peak agglomeration that maximizes the productivity gains³. We find that a firm (with its time invariant idiosyncratic characteristics and for a given level of employment and capital) that would move from a location with no other workers to a location with 650 employees from its own sector (the peak of the observed distribution in France) would gain 25% in TFP. However, going to an "over-crowded" area (with more than 9000 employees) would eliminate these TFP gains. Hence, geography matters a lot for French firms and they are aware of it as they seem to take into account the TFP gains in their location choice. Indeed, when we compare the geography that would maximize productivity gains and the observed geography, we strikingly find very little difference between the two. From this point of view, our paper puts into doubt the rationale of cluster policies which aim is to increase the size of clusters.

The empirical literature on agglomeration economies began in the 1970's with Shefer (1973) and Sveikauskas (1975)'s pioneer works. Rosenthal and Strange (2004) mention in their survey many empirical studies on agglomeration economies. The elasticity of productivity to the size of the city or to the size of the industry generally lies between 3% and 8%, but until the mid 1990's, measures of agglomeration externalities suffered from serious endogeneity problems.

From a technical point of view, the estimation of geographical externalities is subject to two main sources of endogeneity: unobserved heterogeneity and simultaneity bias. Most agglomerated areas might be areas with better endowments (public infrastructure, climate etc.) or may attract more productive firms; agglomeration economies could be overestimated if this unobserved heterogeneity is not taken into account. Moreover, the increase (or decrease) of local employment may be, at least partly, due to cyclical effects which also impact firms' performance; this simultaneity issue could also bias the results.

Another issue is the inadequacy of the data used in many empirical studies to measure directly agglomeration externalities. Indeed, the theories which underlie MAR-type externalities are microeconomic in essence. Consequently, their empirical validity is best verified with firm level data. In the absence of such data, many studies have tried to measure indirectly agglomeration externalities, using more aggregated data on sectoral employment at local level.

This literature is largely inspired by theories on endogenous growth (Romer (1986), Lucas (1988)). The idea is that geographical agglomeration does not only favor productivity but also productivity growth at firm level. Consequently, at local level, the more an area is specialized in a given industry, the higher is productivity growth of that industry in that area. But researchers were confronted to a lack of reliable data on firms' and regions' productivity, whereas data on local employment by industries were available. Therefore, a crucial assumption is made in most studies on dynamic externalities: productivity growth is supposed to imply employment growth. Consequently, whereas the theory invites us to assess the impact of geographic agglomeration of activities on firms' productivity, Glaeser, Kallal, Scheinkman, and Shleifer (1992), Henderson, Kuncoro, and Turner (1995) or Combes (2000) examine the

³Au and Henderson (2006) analyze this question for Chinese cities and also find a bell shaped curve.

effect of specialization and of industrial diversity on local employment growth.

Glaeser, Kallal, Scheinkman, and Shleifer (1992) find, on American data, that sectoral employment growth at local level is negatively affected by specialization. On the contrary, industrial diversity seems to favor sectoral employment growth. Combes (2000) also finds a rather negative impact of specialization on employment growth in both industry and services in France. Henderson, Kuncoro, and Turner (1995) show on American data that mature industries tend to be subject to localization economies, but not to urbanization externalities, whereas high-tech industries are subject to both economies.

Therefore, most studies based on employment growth do not confirm the existence of marshallian externalities, and even tend to show a negative impact of specialization on local sectoral employment growth, whereas evidence about urbanization economies is rather mixed.

Nevertheless, Cingano and Schivardi (2004) cast serious doubt on those results. Indeed, the assumption that productivity growth implies employment growth is not valid for example if the demand for the good is too inelastic. In this case, localization economies may both enhance productivity and reduce local employment. The authors confirm this intuition by studying the Italian case. They use firm level data to construct indices of Total Factor Productivity at industry and Local Labor Systems⁴ level. They find a positive and significant impact of specialization and city size, but no effect of diversity and competition. An interesting finding is that the same regression realized on local sectoral employment growth rate gives results very similar to Combes' or Glaser and al.'s ones: specialization and city size affect negatively local employment growth, whereas sectoral diversity has a positive impact. Consequently, the conclusions about agglomeration economies based on the local sectoral employment growth approach appear spurious.

Theories about agglomeration originally analyzed the impact of local industrial structure on productivity level. Ciccone and Hall (1996) study the impact of county employment density on American states' labor productivity. Their work is the first to address directly and carefully the endogeneity issues we mentioned above. The authors insist on the fact that if there are unmeasured and/or unobserved differences in the determinants of productivity across states, and if these determinants are correlated with states density, the measure of the returns to density by simple OLS may be spurious. They take the example of climate or transportation infrastructures which will both enhance workers' productivity and the attractiveness of the place. They consequently resort to an instrumental variables approach. Also controlling for the average level of education within the state or the county, the authors find that a doubling of local employment density increases labor productivity by 5% to 6%.

Ciccone and Hall's article represents an important step in the empirical approach of agglomeration externalities. Nevertheless, their work still relies on an aggregate measure of labor productivity, which has two consequences:

- given that agglomeration externalities should affect productivity at the firm level, testing such mechanisms without firm level data can raise doubts on the exercise.
- agglomeration externalities could impact firms' TFP and not only labor productivity. This could bias the results.

In the present paper, the use of firm panel data allows a careful treatment of endogeneity issues and a measurement of agglomeration externalities which is very close to the micro

⁴Defined on the basis of workers' commuting.

theories. As far as we know, Henderson (2003) is the only paper who uses plant level data for such an analysis and is the closest to the present paper. His data is available at five years intervals from 1972 to 1992. He estimates a plant level production function for two broad sectors, machinery industries and high-tech industries, and measures the elasticity of TFP to the number of other plants of the same industry in the county. Using industry-time and plant/location fixed effects, he finds a positive and significant elasticity of 8% in the high tech industry only. He does not find evidence of urbanization economies. The use of fixed effects accounts for a large part of unobserved heterogeneity. Henderson also addresses the question of simultaneity bias by adding location-time fixed effects.

Our paper goes further than the study by Henderson in several directions. We use French firms panel data, with yearly observations from 1996 to 2004. Our sample is larger and more complete than Henderson's one which allows us to deal with simultaneity bias and instrumentation more directly. Our sample covers the whole manufacturing sector. We adopt a production function framework and we decompose carefully the agglomeration effects into own industry/other industries externalities, diversity and competition effects. We also address the issue of spatial selection of firms. Finally, using a non linear specification, we can describe the geography that maximizes productivity gains from clustering and compare it to the observed geography.

Section 2 details our empirical strategy, section 3 then proceeds to a description of the data we use, while section 4 presents basic results and section 5 makes a certain number of robustness checks.

2 Estimating agglomeration externalities: empirical strategy

2.1 The model

Through the different channels mentioned above, agglomeration economies are generally assumed to improve total factor productivity (TFP) of firms. When firm-level data is available, this suggests a natural empirical strategy, based on the estimation of a Cobb-Douglas production function:

$$Y_{it} = A_{it} K_{it}^{\alpha} L_{it}^{\beta} \quad (1)$$

where Y_{it} is value-added of firm i at time t , A_{it} is TFP, K_{it} the capital stock and L_{it} the labor-force (in terms of employees) of firm i at time t . We then assume that TFP of firm i depends upon a firm-level component, U_{it} , but also on its immediate environment in terms of localization and urbanization economies:

$$A_{it} = (\text{LOC}_{it}^{sz})^{\delta} (\text{URB}_{it}^{sz})^{\gamma} U_{it}, \quad (2)$$

where LOC_{it}^{sz} is a measure of localization economies and URB_{it}^{sz} is a measure of urbanization economies for firm i , which belongs to sector s and area z , at time t . Log-linearizing the production function, one obtains:

$$y_{it} = \alpha k_{it} + \beta l_{it} + \delta \text{loc}_{it}^{sz} + \gamma \text{urb}_{it}^{sz} + u_{it}, \quad (3)$$

where lower case denotes the log of variables in equations (1) and (2). The model can be estimated by a simple Ordinary Least Squares (OLS) regression if all the independent variables are observable and at least weakly exogenous, but this hypothesis is rarely respected. Consequently, several estimation issues arise that we now detail.

2.2 Estimation issues

Two main issues arise when estimating such a production function: unobserved heterogeneity and simultaneity. In this subsection, we successively analyze those difficulties and ways to solve them.

2.2.1 Unobserved heterogeneity

Some characteristics, unobserved by the econometrician, can both be related to the value added of the firm and to some explanatory variables. In this case, u_{it} is correlated with the independent variables; consequently, the OLS estimations of the coefficients are potentially biased, since the endogenous variables will partly capture the effect of unobserved characteristics. This issue is better known as the “unobserved heterogeneity” problem. In our specification, k_{it} , l_{it} , loc_{it}^{sz} and urb_{it}^{sz} are all likely to be correlated with u_{it} :

- As an example, if an entrepreneur is less risk-averse than the others, he might tend to distort its labor-capital mix in a particular way, have different innovation strategies and also might tend to seek more risky (and potentially more lucrative) markets. Not taking into account firm’s invariant characteristics potentially biases the estimation of α and β .
- Local climate, transportation infrastructures, natural resources or public services to firms can in many ways increase the value-added of a firm. In the same time, a region richly endowed with those environmental elements will be more attractive for firms. There is a positive correlation between unobserved (or unmeasured) firm’s environmental variables and localization and/or urbanization indices which again potentially biases the estimation of δ and γ .

The first estimations of agglomeration economies were often based on aggregate and cross-sectional data (as Shefer (1973) for example) that could not take into account the potential biases just mentioned. The use of individual panel data enables us to address directly these questions.

If we consider firms which do not change industry or region across time, the firm-level and environmental unobserved characteristics mentioned can be appropriately dealt with using firms’ fixed effects which will take into account all firms’ specific characteristics that are invariant across time, whether or not those characteristics are observable. This amounts to assuming that $u_{it} = \phi_i + \epsilon_{it}$:

$$y_{it} = \alpha k_{it} + \beta l_{it} + \delta \text{loc}_{it}^{sz} + \gamma \text{urb}_{it}^{sz} + \phi_i + \epsilon_{it}, \quad (4)$$

where the remaining error term ϵ_{it} is now assumed to have the required properties, and in particular no to be correlated with explanatory variables.

Combes, Duranton, and Gobillon (2007) and Combes, Duranton, Gobillon, and Roux (2008) have shown the spatial sorting of workers to be important. That spatial sorting must be reflected in firms’ TFP but we do not have information about the skills of workers that firms employ. If skills composition of firms’ workforce does not change over the period, individual fixed effect will also take into account the heterogenous quality of labor among firms. Note that doing so, we nevertheless might eliminate a part of the effect we want to measure. Indeed,

if the temporal scope of agglomeration economies is quite long and if firms “capitalize” year after year those externalities, using an individual fixed effect purges our measure of firms’ TFP from part of the gains. Consequently, the elasticity we obtain should be seen as a lower bound of those gains.

Using a panel of firms over several years, one can use standard fixed effects techniques, which involve the introduction of a set of firm dummies, or equivalently mean-differencing of (4). Alternatively, one can “eliminate” ϕ_i using a time differencing approach. The estimated equation is in this case:

$$\Delta y_{it} = \alpha \Delta k_{it} + \beta \Delta l_{it} + \delta \Delta \text{loc}_{it}^{sz} + \gamma \Delta \text{urb}_{it}^{sz} + \Delta \epsilon_{it} \quad (5)$$

We use both methods in our study and not surprisingly, we obtain very similar results (see subsection 4.2). Unobserved heterogeneity is not however the only source of endogeneity affecting agglomeration effects estimation.

2.2.2 Simultaneity bias

Estimating agglomeration economies with a production function approach raises simultaneity issues:

- at the beginning of the year, the entrepreneur can anticipate a positive (or a negative) economic shock and consequently decide of the amount of capital and labor he will use. From an econometric point of view, there is in equation (4) a possible correlation between ϵ_{it} , k_{it} and l_{it} so that the estimation of α and β is potentially biased.
- as a consequence of the negative (or positive) economic shock in the region or in the industry, other firms may close (open) or lay off (hire) employees. ϵ_{it} , loc_{it}^{sz} and urb_{it}^{sz} are also possibly correlated and the estimations of δ and γ may be spurious.

To address the simultaneity issue, different approaches are possible. Olley and Pakes (1996) developed a semiparametric procedure to solve endogeneity and selection problem, making assumptions on the dynamics of capital accumulation. We prefer a GMM approach. The method follows Bond (2002) and Griliches and Mairesse (1995): we start by taking first-differences of each variable, to address the unobserved heterogeneity issue. We then instrument first-differenced independent variables by their level at time $t - 2$, following a GMM procedure. The underlying econometric assumption is that the idiosyncratic shock at time $t - 2$ is orthogonal to $\Delta \epsilon_{it}$. Under this assumption, the instruments are exogenous⁵.

3 Data and variables

We present here the data we use, the way we build our sample and some issues about the construction of our variables.

⁵We have also ran our regressions with other measures of TFP, such as the one developed by Levinsohn and Petrin (2003), and results remain unchanged

3.1 The French annual business survey

We use French annual business surveys⁶ data, provided by the French ministry of Industry. We have information at firm and plant level. The data set covers firms with more than 20 employees and all the plants of those firms over the 1996-2004 period. There is consequently a selection of firms in our sample according to their size. Theoretical works (Melitz and Ottaviano 2008, Baldwin and Okubo 2006) have shown that there might be spatial selection of firms, the most productive ones being predominantly located in denser areas. Yet, we know that bigger firms are more productive than the others. The incompleteness of our sample could consequently be a problem. But as we measure agglomeration economies at firm level, accounting for individual fixed effects, the absence of firms under 20 employees is not, according to us, too problematic. It would be if there was heterogeneity in the sensitivity to agglomeration economies according to the size of the firm, but we have verified that it was not the case.

At the firm level, we have all balance-sheet data (production, value-added, employment, capital, exports, aggregate wages etc.) and information about firm location, firm industry classification and firm structure (number of plants, etc.). At the plant level, data are less exhaustive; they mainly contain plant location, plant industry classification, number of employees and information about the firm the plant belongs to.

3.2 The variables

Firm value added, employees and capital (measured at the beginning of the year) are directly available in the business annual surveys. The creation of agglomeration variables is more elaborate. First of all, the geographical and the sectoral level of aggregation could have an impact on our measure of agglomeration economies⁷. This is why we decided to focus on two geographical entities, the départements, which are administrative entities (there are 100 départements in France, of which 4 are overseas départements) and the employment areas, which are economic entities defined on the basis of workers' commuting (there are 348 employment areas in metropolitan France). From a sectoral point of view, we consider the French sectoral classification (Naf) at both the three and two-digit levels. Consequently, we create our agglomeration variables at four levels: employment area/Naf 3-digit, employment area/Naf 2-digit, département/Naf 3-digit and département/Naf 2-digit. The definition of our variables follows:

- *localization economies*: to deal with intra-industry externalities, we compute, for each firm, the number of other employees working in the same industry and in the same area. Concretely, we use the annual business surveys at plant level and calculate the number of workers by year, industry and area. For firm i , in industry s , in area z at time t , we then define our localization economies variable as:

$$\text{loc}_{it}^{sz} = \ln(\text{employees}_i^{sz} - \text{employees}_{it}^{sz} + 1)$$

At this stage, two remarks are in order. First, ideally, we should estimate a production function at plant level. But capital data are only available at the firm level, which

⁶Called in French "Enquêtes annuelles d'entreprises".

⁷For more details about the impact of spatial zoning on economic geography estimations, see Briant, Combes, and Lafourcade (2007)

is a problem for multi-plants firms. Rather than making strong assumptions on the distribution of capital among plants, we run our estimations on single-plant firms only. To check that the absence of multi-plants firms does not bias our results, we also run the estimations on a sample with multi-plants firms ; the results remain very similar (see subsection 5.1). When conserving multi-plants firms in the sample, we use for agglomeration variables the values relative to the area where the firm's headquarter is registered. However, a firm can have several plants in different départements or employment areas, so that it can declare a number of employees greater than the number of workers in the industry on its territory of registration; in other words, $(\text{employees}_t^{sz} - \text{employees}_{it}^{sz})$ is possibly negative. In that case, we drop the observation⁸.

- *urbanization economies*: we use two variables to capture urbanization economies. The first one is the number of workers in other industries on the territory z where firm i is located.⁹ Using the same notation, we have:

$$\text{urb}_t^{sz} = \ln(\text{employees}_t^z - \text{employees}_t^{sz} + 1)$$

We also add an Herfindahl index of diversity, defined as follows:

$$H_t^{sz} = \sum_{j \neq s} \left(\frac{\text{employees}_{jt}^z}{\text{employees}_t^z - \text{employees}_t^{sz}} \right)^2$$

The index $\text{div}_t^{sz} = \ln \left(\frac{1}{H_t^{sz}} \right)$ will be an indicator of the diversity that firms of industry s face on territory z at time t .

We introduce a last variable to control for local strength of competitive pressure. The use of such a variable aims to test Porter's idea about competition and agglomeration. According to him, competition whips up innovation so that more intense competition within clusters improves firms' performance ((Porter 1998)). We therefore use an Herfindahl index of industrial concentration:

$$\text{Herf}_t^{sz} = \sum_{i \in S_t^z} \left(\frac{\text{employees}_{it}^{sz}}{\text{employees}_t^{sz}} \right)^2$$

where S_t^z is the set of firms belonging to industry s on territory z at time t ¹⁰. The variable $\text{comp}_t^{sz} = \ln \left(\frac{1}{\text{Herf}_t^{sz}} \right)$ measures the degree of competition a firm of sector s faces on territory z at time t . This gives us the relation we want to bring to data:

$$y_{it} = \alpha k_{it} + \beta l_{it} + \delta \text{loc}_{it}^{sz} + \gamma \text{urb}_{it}^{sz} + \mu \text{div}_t^{sz} + \lambda \text{comp}_t^{sz} + \phi_i + \epsilon_{it}. \quad (6)$$

Note that alternative specifications are sometimes used in the literature; they take specialization and density in the area as proxies for localization and urbanization economies. We will address that issue more in details in subsection 4.3 and show that it does not change much our results.

⁸Since information can circulate among the different plants of a firm, this also makes sense to conserve multi-plants firms in the sample and to make the hypothesis that all the plants potentially benefit from the spillovers originating from the area of the headquarter (in particular from knowledge spillovers).

⁹We can note that from the point of view of the firm, the variables l_{it} , loc_{it}^{sz} and urb_{it}^{sz} operate an exhaustive tripartition of local employment.

¹⁰We constructed Herf_t^{sz} from plant level data, so that employees_{iszt} is really the number of employees working in plants of firm i on territory z at time t .

3.3 Construction of the sample

We create 4 samples, crossing the two territorial levels (départements and employment areas) and the two sectoral classifications (Naf 3-digit and Naf 2-digit) we are interested in.

From a geographical point of view, we drop all firms located in Corsica and in overseas départements. Consequently, our sample covers the 94 and the 341 continental French départements and employment areas respectively. Industry-wise, we keep in the sample firms that belong to manufacturing sectors only. Firms in the food-processing sector have been dropped, since the information related to those comes from a different survey, not entirely compatible with the rest of manufacturing. The sample we use in our estimations spans over nineteen 2-digit and eighty-eight 3-digit industrial sectors¹¹.

For each sample, we drop all firms which changed geographical unit or industrial sector during the period. We conserved single-plant firms only and we also made basic error checks; among other things, we dropped all observations for which value-added, employment or capital were missing, negative or null. We deflated value-added data by an industry-level price index and capital data by a national investment price index.

Finally we cleaned up our sample from large outliers, dropping the 1% extreme values for the following variables: mean work productivity, capital intensity, yearly capital growth rate, yearly employment growth rate, yearly mean work productivity growth rate, yearly mean capital intensity growth rate.

3.4 Summary statistics

In this section, we present summary statistics for the Employment area/Naf 3-digit sample.

Table 1 shows how our sample exhibits temporal attrition. This is due to the fact that during the recent period, manufacturing industry has been losing, in France as in other industrial countries, many firms and employees.

Table 1: Temporal composition of the sample Naf 3-digit/Employment area

| Year | Observations | Percent | Cum. Percent |
|-------|--------------|---------|--------------|
| 1996 | 11555 | 12.22 | 12.22 |
| 1997 | 10918 | 11.54 | 23.76 |
| 1998 | 10895 | 11.52 | 35.28 |
| 1999 | 10710 | 11.32 | 46.61 |
| 2000 | 10513 | 11.12 | 57.72 |
| 2001 | 10403 | 11.00 | 68.72 |
| 2002 | 10312 | 10.90 | 79.63 |
| 2003 | 9883 | 10.45 | 90.08 |
| 2004 | 9384 | 9.92 | 100.00 |
| Total | 94573 | 100.00 | |

Table 2 shows the usual descriptive statistics of our variables. First note that most variables exhibit strong variability, as shown by the large values of standard-deviations respective to their mean.

As our data source mainly covers firms with more than 20 employees, the average size of the single-plant firms of our sample is quite large (63 employees). The minimum value for the

¹¹In the French 2-digit classification, manufacturing sectors correspond to sector 17 to sector 36, sector 23 excluded.

localization economies variable (in terms of employees and of firms) is zero: some firms are the sole representative of their industry in their employment area. For those firms, there are consequently no localization economies¹².

Table 2: Summary statistics Naf 3-digit/Employment area

| Variable | Observations | Mean | Std. Dev. | Min | Max |
|--|--------------|----------|-----------|-------|----------|
| Value added | 94573 | 2625.15 | 5897.22 | 32.39 | 413909.9 |
| Firm's employment | 94573 | 63.34 | 100.88 | 1 | 6616 |
| Firm's capital | 94573 | 2799.35 | 11154.95 | 7.30 | 1052349 |
| Firm's capital intensity | 94573 | 34.05 | 33.81 | 0.83 | 258.64 |
| Firm's labor productivity | 94573 | 39.94 | 18.53 | 11.68 | 161.63 |
| # employees, other firms, same industry-area | 94573 | 1098.82 | 2817.19 | 0 | 24475.01 |
| # other firms, same industry-area | 94573 | 18.49 | 48.76 | 0 | 520 |
| # other employees, same area | 94573 | 19885.89 | 24433.19 | 16 | 115785 |
| # other firms, same area | 94573 | 275.35 | 376.74 | 3 | 2164 |

Value-added, capital, capital intensity and labor productivity are expressed in thousands of real euros

Note that summary statistics are basically the same in all samples except for agglomeration variables. As expected, localization variables are bigger when measured at the level of the département rather than at the employment area level (this is also true for urbanization variables), and at the Naf 2-digit level rather than at the Naf 3-digit level (whereas urbanization variables are bigger at Naf 3-digit level). We can also note that the minimum value of firm's number of employees is 1 and not 20; indeed, under certain conditions, firms below 20 employees can be counted in the annual business surveys, but they are very few.

4 How large are agglomeration economies ?

As analysed in subsection 2.2, estimations of production functions suffer from two main biases, unobserved heterogeneity and simultaneity. We address those two problems through a fixed effects approach first, and then through a GMM approach.

4.1 Measuring agglomeration economies taking into account unobserved heterogeneity

All the variables in our regression are potentially correlated with omitted time-invariant variables (see subsection 2). To capture these, we gradually add fixed effects to the simple OLS regression. To capture shocks which affected all firms of the sample in a given year, we also use year fixed effects. Results are presented in table 3.

The first four regressions ignore competition and diversity effects.

According to the simple OLS regression of column (1), increasing by 10% the number of other workers of the same industry-area, keeping the size of the other sectors in the area constant, increases the value added of a firm by 0.05%. On the other hand, increasing the size of the other sectors in the area by 10%, increases the value added of a firm, all else equal, by 0.71%. Those results would indicate a domination of urbanization economies at firm level.

¹²Since $\text{loc}_{it}^{sz} = \ln(\text{employees}_{it}^{sz} - \text{employees}_{it}^{sz} + 1)$, $\text{loc}_{it}^{sz} = 0$ when $\text{employees}_{it}^{sz} - \text{employees}_{it}^{sz} = 0$.

Table 3: Fixed effects approach, Naf 3-digit/Employment Area

| Model : | Dependent Variable: ln value added | | | | | | | |
|---|------------------------------------|-------------------------------|-------------------------------|-------------------------------|--------------------------------|--------------------------------|--------------------------------|-------------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| ln employees | 0.795 ^a (0.005) | 0.806 ^a (0.005) | 0.817 ^a (0.005) | 0.753 ^a (0.008) | 0.795 ^a (0.005) | 0.803 ^a (0.005) | 0.817 ^a (0.005) | 0.753 ^a (0.008) |
| ln capital | 0.158 ^a (0.003) | 0.159 ^a (0.003) | 0.141 ^a (0.003) | 0.078 ^a (0.004) | 0.159 ^a (0.003) | 0.159 ^a (0.003) | 0.141 ^a (0.003) | 0.078 ^a (0.004) |
| ln (# employees, other firms, same industry-area+1) | 0.005 ^a (0.001) | 0.001 (0.001) | 0.002 (0.001) | 0.007 ^a (0.002) | 0.005 ^a (0.002) | 0.008 ^a (0.002) | 0.002 (0.002) | 0.005 ^b (0.002) |
| ln(# employees, other industries, same area+1) | 0.071 ^a (0.003) | -0.008 (0.020) | -0.005 (0.019) | 0.025 (0.016) | 0.076 ^a (0.003) | 0.014 (0.020) | -0.036 ^c (0.019) | 0.024 (0.016) |
| ln competition | | | | | -0.001 (0.004) | -0.024 ^a (0.004) | -0.003 (0.005) | 0.009 (0.006) |
| ln sectoral diversity | | | | | -0.018 ^a (0.005) | 0.031 ^b (0.014) | -0.032 ^b (0.013) | 0.002 (0.010) |
| Time fixed effect | yes | yes | yes | yes | yes | yes | yes | yes |
| Employment area fixed effects | no | yes | yes | no | no | yes | yes | no |
| Industry fixed effects | no | no | yes | no | no | no | yes | no |
| Firm fixed effects | no | no | no | yes | no | no | no | yes |
| N | 94573 | 94573 | 94573 | 94573 | 94573 | 94573 | 94573 | 94573 |
| R ² | 0.789 | 0.810 | 0.833 | 0.421 | 0.789 | 0.810 | 0.833 | 0.421 |

Note: Standard errors in parentheses. ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels. Standard errors are corrected to take into account individual autocorrelation.

But we saw that the estimation of agglomeration economies had to deal with spatial selection of firms. Several types mechanisms may be at work:

1. There may be better areas than others in terms of transportation infrastructures, climate, technology and skill availability, and clusters form in those regions rather than others, which creates a positive correlation between unobserved location attributes and the localization variable.
2. Furthermore, it is likely that firms which are most clustered are also those with better individual characteristics. One explanation could be a self-selection process à la Melitz and Ottaviano (2008). Baldwin and Okubo (2006) marry Melitz and Ottaviano's approach with a New Economic Geography model and show how one could overestimate empirically agglomeration economies if selection issues are not properly taken into account : since the most productive firms also sell more, they are more likely to gain from forward and backward linkages in clusters; moreover, only the most productive firms find it profitable to locate in densely populated clusters, where they have to face higher competition in order to reap the agglomeration benefits of the area.

Regression (2) reinforces the first of the two arguments just made. Indeed, in regression (1), urbanization economies seem clearly more important than localization externalities. But when employment areas fixed effects are added, the coefficients on localization economies variable and on the size of the other industries decrease drastically and become insignificant. This points to an explanation where "good places" attract firms in all industries. Once the attractiveness of those places is controlled for by fixed effects in column (2), agglomeration variables are not significant any more. The same type of analysis can be made with regressions (5) and (6), when competition and sectoral diversity variables are controlled for, but now, localization variable resists to the introduction of area fixed effects.

On the other hand, no clear message emerges for the coefficient on competition within the industry-area. The addition of area-fixed effects in regression (6) reduces the coefficient on competition, which could corroborate the Baldwin-Okubo selection mechanism since area-fixed effects control for the mean productivity of firms in the area. The coefficient is insignificant in regression (8), with firms-fixed effects, which casts doubt on the pro-competitive effect à la Porter.

Results show that localization economies affect positively productivity after controlling for firm-level heterogeneity. Urbanization economies, sectoral diversity and competition have no influence. This confirms that taking into account individual unobserved heterogeneity is extremely important¹³.

It can also be noted that the primary inputs' coefficients decrease when introducing fixed effects in the regression, maybe because fixed effects capture some kind of heterogeneity in the quality of inputs¹⁴. It is particularly striking for capital, which coefficient is around 0.07 once firms fixed effects have been taken into account. This coefficient seems abnormally low, but it is in fact quite usual in the literature on production functions (see Bond (2002) and Griliches and Mairesse (1995)). It can be explained by a measurement error problem (since

¹³This is technically confirmed by Hausman tests.

¹⁴If more productive firms are also firms with better inputs, simple OLS measure an inflated positive correlation between value-added and inputs.

capital is frequently badly measured) and a problem of simultaneity. This is why our first results should be refined with an instrumental variables approach.

4.2 Measuring agglomeration economies taking into account both unobserved heterogeneity and simultaneity

All our variables are potentially subject to a simultaneity bias (see subsection 2.2.2). In order to correct it, and to reduce the measurement error problem on capital data, we resort to an instrumental variables approach. The procedure first takes the first difference of all variables (to eliminate the individual fixed effect) and then instruments each first-differenced independent variable by its level at time $t - 2$. We also add to the set of instruments variables such as loc_{it}^{sz} and urb_{it}^{sz} at time $t - 2$ expressed in terms of plants and firms, and the total number of employees, firms and plants in the employment area at time $t - 2$. The underlying assumption is that shocks are not anticipated by agents so that each variable at time $t - 2$ is orthogonal to $(\epsilon_t - \epsilon_{t-1})$. Such a method reduces drastically the size of the sample, since an observation participates to the estimation if and only if, for the same firm, the two preceding observations are also available. Consequently, the first two years of the sample, 1996 and 1997, are, among others, automatically removed from the sample. Results are presented in table 4.

Regressions (2) and (6) are instrumented regressions and (3) and (7) are GMM regressions, where we take into account autocorrelation at the firm level. Regressions (4) and (8) are GMM regressions where standard errors are corrected using Moulton's method. Moulton (1990) showed that regressing individual variables on aggregate variables could induce a downward bias in the estimation of standard-errors. Here, the aggregate variables are $\ln(\# \text{ employees, other industries, same area}+1)$, $\ln(\text{competition})$ and $\ln(\text{sectoral diversity})$; they are time-industry-employment area variables. Consequently, Moulton's correction consists in clustering the observations at time-industry-employment area level. Our specification is robust to the Sargan-Hansen test of joint validity of instruments.

As expected, inputs are highly significant; but it is interesting to note that our method boosts the coefficient on capital to a more reasonable level (0.217, which is consistent with the results in the literature on production functions). Note that not surprisingly, the coefficient on own employees is quite large : single-plant firms are more labor-intensive than multi-plant ones, which is a well-known result. In subsection 5.1, we show that on the whole sample, French firms exhibit a production function which is close to constant returns to scale.

Our results show that there exist positive and significant localization economies: for a firm, all other things being equal, a 10% increase the number of workers of the industry in the rest of the employment area increases the value added produced by that firm by around 0.4-0.5%. The number of employees in the other sectors of the area, competition and sectoral diversity have no significant impact.

While we expected a positive correlation between shocks and the agglomeration variables loc_{it}^{sz} and urb_{it}^{sz} , our results show that it is not the case.

To sum up, for French firms, there is no evidence of Jacob's urbanization economies: ceteris paribus, sectoral diversity and the scale of activities in other sectors have no significant effect on firms' TFP. The only source of agglomeration economies are MAR-type externalities, with a positive significant coefficient indicating that a 10% increase of employment in neighboring firms of the same industry increases a firm productivity by around 0.4-0.5%.

Table 4: Instrumental variables approach, Naf 3-digit/ Employment Area

| Model : | Dependent Variable: $\Delta \ln(\text{value added})$ | | | | | | | |
|---|--|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\Delta \ln(\text{employees})$ | 0.512 ^a (0.009) | 0.894 ^a (0.077) | 0.896 ^a (0.077) | 0.896 ^a (0.099) | 0.512 ^a (0.009) | 0.898 ^a (0.083) | 0.90 ^a (0.082) | 0.898 ^a (0.106) |
| $\Delta \ln(\text{capital})$ | 0.071 ^a (0.006) | 0.214 ^a (0.024) | 0.215 ^a (0.024) | 0.216 ^a (0.031) | 0.071 ^a (0.006) | 0.211 ^a (0.024) | 0.212 ^a (0.024) | 0.212 ^a (0.030) |
| $\Delta \ln(\# \text{ employees, other firms, same industry-area+1})$ | 0.006 ^a (0.002) | 0.050 ^a (0.019) | 0.051 ^a (0.019) | 0.051 ^b (0.024) | 0.006 ^a (0.002) | 0.042 ^b (0.020) | 0.043 ^b (0.020) | 0.043 ^c (0.025) |
| $\Delta \ln(\# \text{ employees, other industries, same area+1})$ | 0.012 (0.015) | -0.013 (0.113) | -0.015 (0.113) | -0.016 (0.157) | 0.016 (0.015) | 0.101 (0.140) | 0.098 (0.140) | 0.101 (0.195) |
| $\Delta \ln(\text{competition})$ | | | | | -0.001 (0.005) | 0.039 (0.034) | 0.045 (0.033) | 0.039 (0.047) |
| $\Delta \ln(\text{sectoral diversity})$ | | | | | 0.011 (0.009) | -0.098 (0.091) | -0.098 (0.091) | -0.092 (0.129) |
| Sargen-Hansen test/p-value | | | | 0.738 | | | | 0.798 |
| N | 54991 | 54991 | 54991 | 54991 | 54991 | 54991 | 54991 | 54991 |
| R ² | 0.123 | 0.024 | 0.023 | 0.022 | 0.123 | 0.019 | 0.017 | 0.019 |

Note: Standard errors in parentheses. ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels.

(1) and (5) simple OLS, (2) and (6) are IV, with standard errors taking into account individual auto-correlation, (3) and (7) are GMM, with standard errors taking into account individual auto-correlation, (4) and (8) are GMM with Moulton standard errors.

4.3 Alternative specification

Some alternative specifications of agglomeration economies are possible. In particular, Combes, Duranton, and Gobillon (2007) and Combes, Duranton, Gobillon, and Roux (2008) adopt a framework where they test the impact on firm’s productivity (or wages) of density and specialization.

We present in that subsection the results of such a specification on our data.

The specialization and density variables are defined as follows:

$$\text{specialization}_{it}^{sz} = \ln \left(\frac{\text{employees}_t^{sz} - \text{employees}_{it}^{sz} + 1}{\text{employees}_t^z - \text{employees}_{it}^{sz} + 1} \right)$$

$$\text{density}_t^{sz} = \ln \left(\frac{\text{employees}_t^z - \text{employees}_t^{sz} + 1}{\text{size of the area, in sq. kms}} \right)$$

As we can see in table 5, results are very similar. The only difference with the former specification is that the interpretation of the localization economies variable must now be interpreted as the impact of the number of employees in firm’s own industry, keeping the size of the area in terms of employees constant. This means that in parallel to the increase of the number of own industry employees, a decrease of the number of employees in other industries must occur. In the empirical literature, these two different specification exist. Here, we follow Henderson (2003) specification.

4.4 Marginal effects and explanatory power of localization economies

In this subsection, we analyse the impact of the choice of classification on the intensity of localization economies; we then study the explanatory power of MAR externalities with respect to firm’s employees and capital.

4.4.1 Different intensities for localization economies or Modifiable Areal Unit Problem

We reproduce the same analysis for the other three levels of sectoral and geographical aggregation. The results of GMM regressions are presented in the appendix. The findings on the evolution and significance of coefficients remain roughly unchanged. However, as we can observe in table 6, the impact of a doubling of localization variable¹⁵ on productivity varies according to the aggregation level. For a given industrial classification, localization economies seem more intense at the département level than at the employment area level. Two explanations are possible: MAR externalities are really greater at département than at employment area level, or the different intensities are only due to statistical noise due to the choice of spatial unit (this problem is also known as Modifiable Areal Unit Problem (MAUP), see Briant, Combes, and Lafourcade (2007)). At this stage, we cannot distinguish between those two effects.

The same observation can be made at a sectoral level. Here again, two interpretations are possible: either it is a simple statistical artefact, or it suggests that Marshall’s externalities involve bigger industrial sectors than Naf 3-digit ones. An explanation can also be found in the existence of vertical externalities along the supply chain, since Naf 2-digit sectors certainly embrace subsectors related by consumer/supplier relationships.

¹⁵If $\ln y = \alpha \ln x$, y increases in percentage by $(2^\alpha - 1) \times 100$ when x is doubling.

Table 5: Instrumental variables approach, Naf 3-digit/Employment Area

| Model : | Dependent Variable: $\Delta \ln(\text{value added})$ | | | | | | | |
|---|--|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\Delta \ln(\text{employees})$ | 0.512 ^a (0.009) | 0.953 ^a (0.084) | 0.955 ^a (0.084) | 0.950 ^a (0.108) | 0.512 ^a (0.009) | 0.940 ^a (0.085) | 0.943 ^a (0.085) | 0.940 ^a (0.109) |
| $\Delta \ln(\text{capital})$ | 0.071 ^a (0.006) | 0.228 ^a (0.025) | 0.229 ^a (0.025) | 0.228 ^a (0.032) | 0.071 ^a (0.006) | 0.227 ^a (0.025) | 0.228 ^a (0.025) | 0.227 ^a (0.031) |
| $\Delta \ln(\text{specialisation})$ | 0.006 ^a (0.002) | 0.060 ^a (0.021) | 0.061 ^a (0.021) | 0.059 ^b (0.027) | 0.006 ^a (0.002) | 0.051 ^b (0.021) | 0.052 ^b (0.021) | 0.051 ^c (0.027) |
| $\Delta \ln(\text{density})$ | 0.024 (0.016) | -0.040 (0.086) | -0.048 (0.085) | -0.044 (0.117) | 0.028 ^c (0.016) | -0.041 (0.092) | -0.040 (0.092) | -0.035 (0.127) |
| $\Delta \ln(\text{competition})$ | | | | | -0.001 (0.005) | 0.054 (0.034) | 0.055 (0.034) | 0.053 (0.048) |
| $\Delta \ln(\text{sectoral diversity})$ | | | | | 0.011 (0.009) | -0.027 (0.082) | -0.028 (0.082) | -0.030 (0.117) |
| Sargan-Hansen test/p-value | | | | 0.713 | | | | 0.832 |
| N | 54991 | 54991 | 54991 | 54991 | 54991 | 54991 | 54991 | 54991 |
| R ² | 0.123 | 0.007 | 0.009 | 0.006 | 0.123 | 0.005 | 0.007 | 0.005 |

Note: Standard errors in parentheses. ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels. (1) and (5) simple OLS, (2) and (6) are IV, with standard errors taking into account individual auto-correlation, (3) and (7) are GMM, with standard errors taking into account individual auto-correlation, (4) and (8) are GMM with Moulton standard errors.

Table 6: Localization economies marginal effect

| Employment area/Naf 3 | Département/Naf 3 | Employment area/Naf 2 | Département/Naf 2 |
|-----------------------|-------------------|-----------------------|-------------------|
| 3.03% | 4.75% | 3.89% | 13.60% |

4.4.2 Explanatory power of localization externalities

The explanatory power of a variable depends on the value of the coefficient attached to it, and on its variability. If a variable has a very low variance, its explanatory power will be weak, even if it has a high coefficient. The explanatory power of an independent variable is strong if, all other things being equal, a standard-deviation of that variable implies a large variation of the dependent variable¹⁶. We consequently calculated the explanatory power of employees, capital and localization variables. The results are presented in table 7.

Table 7: Explanatory power of main variables

| Variable | Employment area/Naf 3 | Département/Naf 3 | Employment area/Naf 2 | Département/Naf 2 |
|--|-----------------------|-------------------|-----------------------|-------------------|
| employees | 119.94% | 149.58% | 160.71% | 174.67% |
| capital | 36.04% | 35.50% | 43.72% | 46.01% |
| # employees, other firms, same industry-area | 5.52% | 6.85% | 5.70% | 14.70% |

Note: The table reads as follows: for a firm, all other things being equal, a standard-deviation with respect to the mean of the number of own employees generates, at Naf 3-digit/Employment area level, an increase of value-added by 119.94%.

When compared to inputs variables, it appears non surprisingly that localization externalities are a second-order determinant of firms' value added. Their impact (between 5 and 15% increase in productivity), without being null is relatively small. Moreover, MAR externalities are again more substantial at the Naf 2-digit level than at the Naf 3-digit.

5 Robustness checks and further issues

5.1 Localization economies and multi-plants firms

Focusing on single-plant firms had two main interests. From a methodological point of view, we saw that the natural level of estimation is the plant level but that data do not allow to do this easily. Estimating production functions at the firm level might be problematic for the definition of multi-plant firms' agglomeration variables: a firm which has plants in several employment areas or départements will be related to its headquarter's agglomeration variables whereas its productive plants may be in other -remote- places. On the other hand, one could think that information circulates among the plants of a given firm, so that spillovers, especially knowledge spillovers, could diffuse within the firm, wherever its plants are.

¹⁶If $\ln y = \alpha \ln x$, we define the explanatory power of x as $[\exp(\alpha \ln(1 + \frac{\sigma_x}{\bar{x}})) - 1] \times 100 = [(1 + \frac{\sigma_x}{\bar{x}})^\alpha - 1] \times 100$, where σ_x and \bar{x} are respectively the standard deviation and the mean of x .

Another more substantive aspect, is that we might conjecture that single-plant firms are more dependent on agglomeration economies than multi-plant ones, since the latter can exploit internal networks that single-plant firms simply do not have. Henderson (2003) investigates this issue and shows that American single-plant firms do not seem to benefit more than multi-plant ones from static externalities, but that they do from dynamic externalities.¹⁷

We therefore estimate our production function on the whole sample of firms, to check how the results are affected by the inclusion of multi-plants firms.

Table 8 presents GMM regressions for this sample.

Concerning environmental variables, results have very similar patterns as before : sectoral diversity and the scale of other industries have no impact, whereas localization economies are positive and almost always significant. Their magnitude is very similar to what was found in section 4.4.1. A notable difference exists for competition variable which is now significant. That result can be the sign of a real difference between single-plant and multi-plant firms. It could also mean that an index of sectoral competition at local level has no sense for firms which have productive plants in several areas.

For the explanatory power of variables, the comment is also roughly the same as in sub-section 4.4.2, and there is no evidence of a lesser dependence on agglomeration economies for multi-plants firms.

The methodological problem involved by multi-plants firms does not seem to be in practice as severe as might be expected.

Moreover, we tested, in unreported investigations, the existence of dynamic externalities. We introduced one-year lagged agglomeration variables in the former specification. We also tested a “real” dynamic model, *à la* Combes, Magnac, and Robin (2004), adding as an independent variable the one-year lagged value added. We do not find any sign of dynamic externalities, neither for multi-plant nor for single-plant firms. But it is possible that our data are insufficient to address this question: since we have annual data for a relatively short period, considering lags longer than a year would reduce dramatically the size of our sample, whereas the scope of dynamic externalities is probably longer (Henderson (2003) has five years lags).

5.2 Who generates externalities: firms or employees?

Theory offers several possible channels for MAR-type economies. A notable alternative is whether externalities transit through firms or workers? For a firm, is it the same to have in the neighborhood one firm of the industry with a hundred employees or ten firms, each of them employing ten workers ? The question is important for policy makers interested in clusters; according to the answer, extensive or intensive development strategy will be preferable.

Henderson (2003) finds that plants generate externalities, but not workers. If we consider each plant as a source of knowledge, this result is the sign, according to Henderson, that information spillovers are more important than labor market externalities.

Our results are quite different. For firm i from sector s in area z at time t , we decompose the number of employees in its own industry-area into two components: the number of firms is sector s in area z at time t and the mean size of those firms.¹⁸ Keeping the *number of*

¹⁷Henderson (2003) measures dynamic externalities using past values of the localization economies variable.

¹⁸Indeed, $\ln(\text{employees}_t^{sz} - \text{employees}_{it}^{sz} + 1)$ is equal to $\ln\left(\frac{\text{employees}_t^{sz} - \text{employees}_{it}^{sz} + 1}{\text{firms}_t^{sz}}\right) + \ln \text{firms}_t^{sz}$.

Table 8: Multi-plants firms and agglomeration economies

| Model : | Dependent Variable: $\Delta \ln(\text{value added})$ | | | |
|---|--|-------------------------------|-------------------------------|-------------------------------|
| | Employment area/Naf 3 | Département/Naf 3 | Employment area/Naf 2 | Département/Naf 2 |
| $\Delta \ln(\text{employees})$ | 0.748 ^a (0.107) | 0.847 ^a (0.105) | 0.828 ^a (0.101) | 0.834 ^a (0.112) |
| $\Delta \ln(\text{capital})$ | 0.223 ^a (0.026) | 0.207 ^a (0.024) | 0.226 ^a (0.025) | 0.223 ^a (0.024) |
| $\Delta \ln(\# \text{ employees, other firms, same industry-area}+1)$ | 0.040 ^c (0.023) | 0.076 ^b (0.033) | 0.054 (0.035) | 0.190 ^a (0.066) |
| $\Delta \ln(\# \text{ employees, other industries, same area}+1)$ | 0.073 (0.164) | 0.002 (0.136) | -0.042 (0.161) | -0.115 (0.146) |
| $\Delta \ln(\text{competition})$ | 0.103 ^b (0.043) | 0.090 ^b (0.040) | 0.083 ^c (0.047) | 0.028 (0.048) |
| $\Delta \ln(\text{sectoral diversity})$ | -0.090 (0.112) | -0.132 (0.084) | 0.049 (0.116) | -0.222 (0.145) |
| Sargan-Hansen test/p-value | 0.583 | 0.194 | 0.074 | 0.252 |
| N | 76209 | 86180 | 88458 | 92765 |
| Centered R ² | 0.057 | 0.041 | 0.048 | 0.035 |

Note: standard-errors in parentheses ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels. Standard-errors are Moulton's standard-errors.

Table 9: Single-plant firms/Explanatory power of main variables

| | Employment area/Naf 3 | Département/Naf 3 | Employment area/Naf 2 | Département/Naf 2 |
|--|-----------------------|-------------------|-----------------------|-------------------|
| employees | 123.86% | 154.88% | 167.11% | 177.92% |
| capital | 38.91% | 36.16% | 47.13% | 46.33% |
| # employees, other firms, same industry-area | 5.04% | 7.96% | 5.73% | 16.00% |

Note: The table reads as follows: for a firm, all other things being equal, a standard-deviation with respect to the mean of the number of own employees, in the Naf 3-digit/Employment area regression, an increase of value-added by 123.86%.

firms constant, an increase of the *mean size* of firms generates an increase of the *total number of employees* in the sector. We present in table 10 the results of GMM estimations.

When the number of own industry firms and their mean size are both taken into account, the latter is the only one to be significant. Interestingly enough, coefficients on mean size variable are very close to those on the localization variable in our first specification, but their significance is boosted.

To sum up, the case of French firms indicates that there are no specific externalities we can attribute to firms *per se* but that there are positive and significant externalities linked to the number of employees in surrounding firms. The number of employees in the other firms is a better indicator of the size of the industry a firm faces on its territory than the number of firms. This points to an interpretation under which localization economies are, for a firm, due to the “thickness” of the industry around it. Our results are interesting for policy-makers; they suggest that boosting externalities within clusters involves the promotion of internal growth of existing firms or the attraction of big firms on the territory rather multiplying the number of small firms. Moreover, our results support those of Ellison, Glaeser, and Kerr (2007), who find, on American data, that input-output linkages and labor pooling are -in this order- the two main determinants of industries co-agglomeration. They also find evidence of knowledge spillovers, but to a lesser degree.

5.3 Externalities and distance

Another important topic concerning localization economies is their geographic scope. Assessing the geographic scope of MAR-type economies is a question of primary importance for public policy makers in order to define the geographic perimeter of their action. Different approaches have been used in the literature to measure the geographic scope of localization economies or agglomeration patterns : discrete ones, based on political boundaries (such as (Henderson 2003)) or continuous ones (such as (Duranton and Overman 2006)). For a detailed survey, see Rosenthal and Strange (2004).

The existing literature suggests that MAR-type externalities are very localized. According to Henderson (2003), once industrial thickness in the county is taken into account, the number of plants of the industry in the rest of the MSA has no impact. Duranton and Overman (2006) find that localization patterns of industries in UK take place within a quite small perimeter. Henderson and Arzaghi (2007) study the advertising industry in Manhattan and find that there is extremely rapid spatial decay in the externalities firms benefit from. We address this question on French data, using two methods:

Table 10: Employees, firms and agglomeration economies

| Model: | Dependent Variable: $\Delta \ln(\text{value added})$ | | | |
|---|--|-------------------------------|-------------------------------|-------------------------------|
| | Employment area/Naf 3 | Département/Naf 3 | Employment area/Naf 2 | Département/Naf 2 |
| $\Delta \ln(\text{employees})$ | 0.923 ^a (0.110) | 0.958 ^a (0.101) | 1.020 ^a (0.107) | 1.026 ^a (0.111) |
| $\Delta \ln(\text{capital})$ | 0.222 ^a (0.032) | 0.198 ^a (0.028) | 0.226 ^a (0.035) | 0.197 ^a (0.032) |
| $\Delta (\ln(\text{Mean size of other firms, same industry-area}+1))$ | 0.052 ^c (0.028) | 0.084 ^b (0.039) | 0.094 ^c (0.050) | 0.191 ^b (0.086) |
| $\Delta \ln(\# \text{ other firms, same industry-area}+1)$ | 0.077 (0.074) | 0.115 (0.115) | 0.040 (0.154) | -0.164 (0.299) |
| $\Delta \ln(\# \text{ employees, other industries, same area}+1)$ | 0.058 (0.229) | 0.021 (0.177) | -0.144 (0.224) | -0.012 (0.219) |
| $\Delta \ln(\text{competition})$ | 0.032 (0.054) | 0.034 (0.052) | 0.041 (0.056) | 0.073 (0.088) |
| $\Delta \ln(\text{sectoral diversity})$ | -0.097 (0.145) | -0.121 (0.114) | 0.047 (0.146) | 0.024 (0.244) |
| Sargan-Hansen test/p-value | 0.531 | 0.210 | 0.816 | 0.349 |
| N | 54991 | 61332 | 62305 | 64714 |
| Centered R ² | 0-.000 | 0-.002 | 0.038 | 0.054 |

Note: standard-errors in parentheses ^a, ^b, and ^c respectively denoting significance at the 1%, 5% and 10% levels. Standard-errors are Moulton's standard-errors.

- “*Market potential*”: we construct a variable of localization economies inspired by the literature on market potential (dating back to Harris (1954)); for a given location z , we use the sectoral employment of all other areas weighted by bilateral distance to z . For firms of industry s , in area z at time t , this variable is:

$$\text{mp}_t^{sz} = \ln \left(\sum_{j \neq z} \frac{\text{employees}_t^{sj}}{d_{zj}} + 1 \right) \quad (7)$$

where d_{zj} is the distance in kilometers between locations z and j .

- *Contiguity*: we run the same regressions, replacing the “market potential” localization variable by the sum of localization variables in the Z contiguous areas (employment areas or départements). For firms of industry s , in area z at time t , our variable of contiguous localization economies is:

$$\text{contig}_t^{sz} = \ln \left(\sum_{j \in Z} \text{employees}_t^{sj} + 1 \right) \quad (8)$$

Results with the “market potential” method are presented in table 11 while results with the contiguity method are in table 16 in the appendix. The “market potential” method shows clearly that once the local thickness of the industry has been taken into account, the number of employees in the same industry on other areas weighted by distance has no significant impact. Moreover, the inclusion of own-industry employment in other areas tends to reduce the significancy of localization variable. Conclusions about the other agglomeration variables remain qualitatively unchanged. The contiguity method confirms those results.

Consequently, our results on French data show that Marshall’s externalities are very localized. This would support the idea that beyond geographical proximity, organisational proximity and social interactions are probably important: externalities probably channel through “high quality” individual and business relationships. A very close geographic proximity seems necessary to facilitate mutual knowledge and social interactions.

5.4 Is there enough clustering?

We found that firm productivity increases with clustering. Does this imply that more clustering is always better and that public intervention to increase the size of clusters is justified? In theoretical models, clustering has the characteristic of an externality: firms benefit from the fact that other firms in the same sector decide to choose to locate nearby. These firms do not internalize the productivity benefit they bring to other firms through this location choice. This suggests that the decentralized equilibrium may be characterized by suboptimal clustering that would translate into suboptimal productivity. This is the basic argument (although not always put in these terms) that many proponents of cluster policies (such as Michael Porter) put forward to defend public policies that help foster larger clusters.

However, besides cluster benefits, congestion effects may also exist. These congestion effects could affect the utility of agents (through increased traffic, pollution etc...) which we cannot measure, but could also impact negatively the productivity of firms. In this case, the productivity-cluster relationship would take the form of a bell curve: productivity would

Table 11: Agglomeration externalities and distance/“Market potential”

| Model: | Dependent Variable: $\Delta \ln(\text{value added})$ | | |
|---|--|-------------------------------|-------------------------------|
| | Employment area/Naf 3 | Département/Naf 3 | Employment area/Naf 2 |
| $\Delta \ln(\text{employees})$ | 0.932 ^a (0.133) | 1.047 ^a (0.119) | 0.984 ^a (0.104) |
| $\Delta \ln(\text{capital})$ | 0.203 ^a (0.039) | 0.169 ^a (0.032) | 0.227 ^a (0.028) |
| $\Delta \ln(\# \text{ employees, other firms, same industry-area}+1)$ | 0.048 ^c (0.027) | 0.095 ^b (0.042) | 0.051 (0.038) |
| $\Delta \ln(\# \text{ employees, same industry, all other areas weighted by distance}+1)$ | -0.104 (0.258) | -0.306 (0.230) | 0.087 (0.106) |
| $\Delta \ln(\# \text{ employees, other industries, same area}+1)$ | 0.089 (0.167) | 0.119 (0.149) | -0.102 (0.117) |
| $\Delta \ln(\text{competition})$ | 0.045 (0.048) | 0.049 (0.045) | 0.035 (0.052) |
| $\Delta \ln(\text{sectoral diversity})$ | -0.079 (0.125) | -0.109 (0.106) | 0.096 (0.117) |
| Sargan-Hansen test/p-val | 0.806 | 0.605 | 0.394 |
| N | 54991 | 61332 | 62305 |
| Centered R ² | 0.007 | 0.038 | 0.014 |
| | | | 0-0.007 |

Note: standard-errors in parentheses. ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels. Standard-errors are Moulton's standard-errors.

increase at low levels of clustering and would then decrease with clustering. The relation would be non linear. This would mean that the effect we measured so far was the mean net effect of localization economies and congestion effects over the distribution of localization economies variable in our sample.

To test the existence of such non-linear localization economies, we introduce in the former regression quadratic and cubic terms of localization variable. We still use a GMM estimation on first differenced variables. Hence, the estimation exploits the temporal variability within firms. Standard errors are still computed using Moulton's method. We drop firms for which the localization variable value is below 20 employees. By construction, there is a difference between the two groups of firms for which the localization variable is below and above 20. The reason is that the Annual Business Surveys data set is restricted to firms with more than 20 employees. Hence, when the localization variable is less than 20, this implies that those plants that make up the localization variable necessarily belong to multi-plant firms only. They are therefore fundamentally different from the group of firms that compose the localization variable when this variable is more than 20¹⁹.

Results are presented in table 12; they show statistical significance for all three terms of localization economies at the Employment area/Naf 3-digit and Département/Naf 3-digit levels.

Table 12: Localization economies vs Congestion effects

| Model | Dependent Variable: $\Delta \ln(\text{value added})$ | |
|---|--|--------------------------------|
| | Employment area/Naf 3 | Département/Naf 3 |
| $\Delta \ln(\text{employees})$ | 0.923 ^a (0.114) | 0.960 ^a (0.104) |
| $\Delta \ln(\text{capital})$ | 0.225 ^a (0.032) | 0.198 ^a (0.029) |
| $\Delta \ln(\# \text{ employees, other firms, same industry-area}+1)$ | -0.251 ^b (0.110) | -0.247 ^a (0.086) |
| $\Delta \ln(\# \text{ employees, other firms, same industry-area}+1)^2$ | 0.085 ^b (0.040) | 0.078 ^a (0.028) |
| $\Delta \ln(\# \text{ employees, other firms, same industry-area}+1)^3$ | -0.006 ^c (0.003) | -0.005 ^b (0.002) |
| $\Delta \ln(\# \text{ employees, other industries, same area}+1)$ | 0.107 (0.227) | 0.028 (0.171) |
| $\Delta \ln(\text{competition})$ | 0.031 (0.051) | 0.051 (0.045) |
| $\Delta \ln(\text{sectoral diversity})$ | -0.150 (0.135) | -0.141 (0.108) |
| Sargan-Hansen test/p-value | 0.983 | 0.665 |
| N | 51491 | 60062 |
| Centered R ² | 0-004 | 0.000 |

Note: standard-errors in parentheses ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels. Standard-errors are Moulton's standard-errors.

We graphically represent those results in figures 1 and 2. The dark curve is the estimate of the TFP surplus gained at each level of the localization variable.

Localization economies net effect has the same shape in both cases: an inverted U-shape

¹⁹When we do not drop the firms for which the localization variable is less than 20, the general shape of the relation between the TFP surplus and localization is unchanged and the coefficients are very similar but less significant.

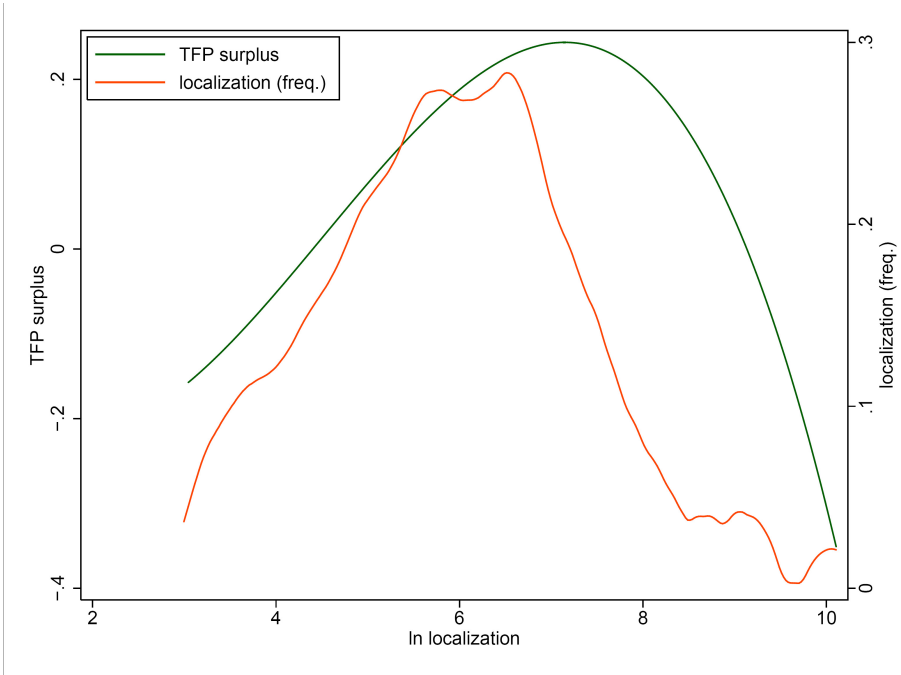


Figure 1: Localization economies - Employment Area/Naf 3-digit

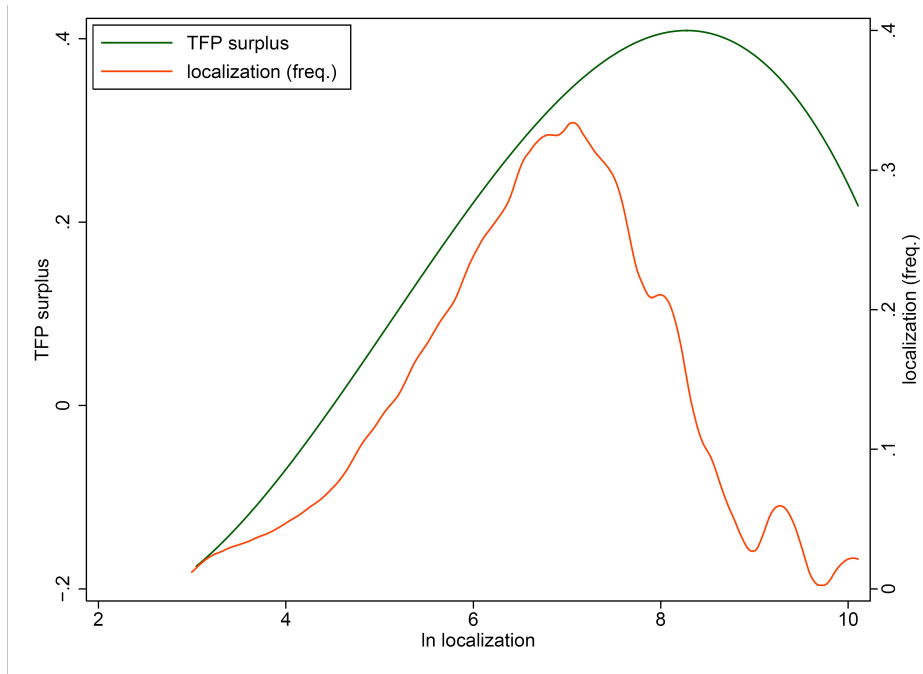


Figure 2: Localization economies - Département/Naf 3-digit

pattern. The net TFP surplus due to localization economies is negative for small values of the localization variables. At the employment area and Naf 3-digit level, the threshold for which the gains from clusters become positive is around 80 employees. Remember this does not include the workers of the firm itself. The second threshold for which the negative effect of cluster dominates the positive effect is around 9250 employees.

The peak, at which the marginal congestion effects of increasing the number of workers in the same employment area and the same sector start to dominate the localization effects, is estimated at 1270 employees.

On the same graph we plot the actual distribution of French firms present in the sample with the grey curve. The peak of the distribution is obtained for firms located in employment areas that have around 650 employees in the same sector (again excluding the workers of the firm itself). Hence, employment areas would tend to be undersized from the point of view of the maximization of firms' TFP. The difference between the observed peak (650) and the estimated one (1270) can seem important. However, the productivity gain for a firm that would go from the former to the latter is small, only 2.1%.

Consequently, the comparison of the two curves suggests that French firms do internalize the productivity gains of clustering when making location choices. Another way to see this is that very few firms (17.7% in our sample) locate in areas for which the TFP surplus that comes from localization is negative.

At the département level, the observed peak is attained for around 1180 employees whereas the estimated peak is attained for around 3920 employees. The productivity gain for a firm that would go from the estimated peak in the distribution to the estimated peak for the TFP gain would now be higher, equal to 6.2%.

We have tried different specifications and this result is robust. The observed distribution of localization variable and the curve of estimated TFP surplus have very similar shape. When differences exist between the two peaks, they all point to the conclusion that clusters are too small, but that the difference in productivity between the two peaks is quite small.

The conclusion that firms internalize productivity gains from clustering is not all that surprising. Our estimation enables us to perform the following thought experiment. Think of a firm (with its time invariant idiosyncratic characteristics and for a given level of employment and capital) that has to choose its location among many employment areas. Strictly speaking, this firm should be small enough so that its location choice does not matter for other firms. Going from an employment area with no other workers from its own sector to an employment area with 650 employees from its own sector (the peak of the observed distribution), the estimated TFP gain is quite large at 25%²⁰. The same gain would be obtained when a firm delocates from a over-crowded area (with 9250 employees of the same sector) to the observed peak of the distribution. This suggests that clusters are a natural implication of firms maximizing profits but that larger clusters are not always better. Hence, one should not conclude from our study that geography does not matter for firms. It matters a lot and firms are aware of it.

²⁰This is consistent with the results of Crozet, Mayer, and Mucchielli (2004) who find that a very important determinant of location choice in France for multinational firms is the localization variable.

6 Conclusion

We have shown that, taking into account many possible biases, localization economies exist in the French economy. A question remains unanswered: who benefits from these productivity gains? Workers, capital owners or land owners? According to Combes, Duranton, and Gobillon (2007), the elasticity of wage to employment's area specialization, on French data, is around 2.1%. Even though the methodology, data and classifications are not strictly comparable, this suggests the returns of localization economies are inferior for wages than those estimated for TFP in our paper, which range between 5 and 10%. This suggests that workers are not able to capture all the gains from localization economies. We also tried to analyze the effect of localization externalities on profits but did not find any conclusive result. This might suggest that a large part of the surplus is captured by the immobile factor, namely land²¹, which would make sense from a theoretical point of view. At this point however, this hypothesis, while plausible, would need further investigation.

Our results have several interesting policy implications in a context in which cluster policies are popular among governments and local authorities. First, the starting point of those who favor cluster policies is right: there are productivity gains to clusters which we measured by localization economies. However, those gains are relatively small and more importantly seem to be already well internalized by firms in their location decisions. The comparison between an estimated geographical distribution of firms that would maximize productivity and the one that is actually observed suggests no large gap, at least in the French case. It points neither to a situation where geography is too concentrated and specialized nor to a geography that needs more clustering. Of course, this result is "only" about productivity and is not about welfare which agglomeration could affect through other channels than through productivity. However, this suggests that even though the starting point of cluster policy advocates is right, their conclusion advocating costly public intervention²² in favor of clusters is not supported by the French evidence.

²¹ This is indeed what Henderson and Arzaghi (2007) find in their study of clusters in advertising in New York city.

²² In Martin, Mayer, and Mayneris (2008), using the same data set as in this paper, we found no evidence that first cluster policy in France, the Systèmes Productifs Locaux, had any effect on productivity of French firms.

References

- AU, C.-C., AND J. V. HENDERSON (2006): "Are Chinese Cities Too Small?," *Review of Economic Studies*, 73(3), 549–576.
- BALDWIN, R. E., AND T. OKUBO (2006): "Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting," *Journal of Economic Geography*, 6(3), 323–346.
- BOND, S. (2002): "Dynamic panel data models: a guide to micro data methods and practice," CEMMAP Working paper CWP09/02.
- BRIANT, A., P. COMBES, AND M. LAFOURCADE (2007): "Do the size and shape of spatial units jeopardize economic geography estimations?," Working paper.
- CICCONE, A., AND R. E. HALL (1996): "Productivity and the Density of Economic Activity," *American Economic Review*, 86(1), 54–70.
- CINGANO, F., AND F. SCHIVARDI (2004): "Identifying the sources of local productivity growth," *Journal of the European Economic Association*, 2, 720–742.
- COMBES, P., G. DURANTON, AND L. GOBILLON (2007): "Spatial wage disparities : sorting matters!," forthcoming in the *Journal of Urban Economics*.
- COMBES, P., G. DURANTON, L. GOBILLON, AND S. ROUX (2008): "Estimating agglomeration economies with history, geology and worker effects," Discussion paper.
- COMBES, P.-P. (2000): "Economic Structure and Local Growth: France, 1984-1993," *Journal of Urban Economics*, 47(3), 329–355.
- COMBES, P.-P., T. MAGNAC, AND J.-M. ROBIN (2004): "The dynamics of local employment in France," *Journal of Urban Economics*, 56(2), 217–243.
- CROZET, M., T. MAYER, AND J.-L. MUCCHIELLI (2004): "How do firms agglomerate? A study of FDI in France," *Regional Science and Urban Economics*, 34(1), 27–54.
- DURANTON, G., P. MARTIN, T. MAYER, AND F. MAYNERIS (2008): "California Dreamin': The Feeble Case for Cluster Policies," Discussion paper.
- DURANTON, G., AND H. G. OVERMAN (2006): "Exploring the Detailed Location Patterns of UK Manufacturing Industries Using Microgeographic Data," Discussion paper.
- ELLISON, G., E. L. GLAESER, AND W. KERR (2007): "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," NBER Working Papers 13068, National Bureau of Economic Research, Inc.
- GLAESER, E. L., H. D. KALLAL, J. A. SCHEINKMAN, AND A. SHLEIFER (1992): "Growth in Cities," *Journal of Political Economy*, 100(6), 1126–52.
- GRILICHES, Z., AND J. MAIRESSE (1995): "Production functions: the search for identification," NBER Working paper No. 5067.
- HARRIS, C. (1954): "The market as a factor in the localization of industry in the United States," *Annals of the Association of American Geographers*, 44(4), 315–348.

- HENDERSON, J. V., AND M. ARZAGHI (2007): “Networking Off Madison Avenue,” forthcoming in the *Review of Economic Studies*.
- HENDERSON, V. (2003): “Marshall’s scale economies,” *Journal of Urban Economics*, 53(1), 1–28.
- HENDERSON, V., A. KUNCORO, AND M. TURNER (1995): “Industrial Development in Cities,” *Journal of Political Economy*, 103(5), 1067–90.
- LEVINSOHN, J., AND A. PETRIN (2003): “Estimating Production Functions Using Inputs to Control for Unobservables,” *Review of Economic Studies*, 70(2), 317–342.
- LUCAS, R. J. (1988): “On the mechanics of economic development,” *Journal of Monetary Economics*, 22(1), 3–42.
- MARSHALL, A. (1890): *Principles of Economics*. Macmillan, London.
- MARTIN, P., T. MAYER, AND F. MAYNERIS (2008): “Public Support to Clusters: A Case study on French Local Productive Systems, 1996-2004,” Discussion paper.
- MELITZ, M. J., AND G. I. P. OTTAVIANO (2008): “Market Size, Trade, and Productivity,” *Review of Economic Studies*, 75(1), 295–316.
- MOULTON, B. R. (1990): “An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit,” *The Review of Economics and Statistics*, 72(2), 334–38.
- OLLEY, S. G., AND A. PAKES (1996): “The dynamics of productivity in the telecommunications equipment industry,” *Econometrica*, 64(6), 1263–1297.
- PORTER, M. (1998): “Clusters and Competition: New Agendas for Companies, Governments, and Institutions,” in *On competition*. Harvard Business School Press.
- ROMER, P. M. (1986): “Increasing Returns and Long-run Growth,” *Journal of Political Economy*, 94(5), 1002–37.
- ROSENTHAL, S. S., AND W. C. STRANGE (2004): “Evidence on the nature and sources of agglomeration economies,” in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson, and J. F. Thisse, vol. 4 of *Handbook of Regional and Urban Economics*, chap. 49, pp. 2119–2171. Elsevier.
- SHEFER, D. (1973): “Localization economies in SMA’s: A production function analysis,” *Journal of Urban Economics*, 13(1), 55–64.
- SVEIKAUSKAS, L. A. (1975): “The productivity of cities,” *Quarterly Journal of Economics*, 89(3), 393–413.

APPENDIX

Table 13: Instrumental variables approach, Naf 3-digit/Département

| Model : | Dependent Variable: $\Delta \ln(\text{value added})$ | | | | | | | |
|---|--|-------------------------------|-------------------------------|-------------------------------|-------------------------------|--------------------------------|-------------------------------|-------------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\Delta \ln(\text{employees})$ | 0.516 ^a (0.009) | 0.922 ^a (0.074) | 0.934 ^a (0.074) | 0.931 ^a (0.094) | 0.516 ^a (0.009) | 0.943 ^a (0.077) | 0.950 ^a (0.077) | 0.951 ^a (0.099) |
| $\Delta \ln(\text{capital})$ | 0.070 ^a (0.006) | 0.186 ^a (0.021) | 0.188 ^a (0.021) | 0.188 ^a (0.027) | 0.070 ^a (0.006) | 0.190 ^a (0.021) | 0.191 ^a (0.021) | 0.192 ^a (0.027) |
| $\Delta \ln(\# \text{ employees, other firms, same industry-area+1})$ | 0.009 ^a (0.003) | 0.054 ^b (0.026) | 0.058 ^b (0.026) | 0.058 ^c (0.033) | 0.008 ^a (0.003) | 0.063 ^b (0.027) | 0.066 ^b (0.027) | 0.067 ^c (0.036) |
| $\Delta \ln(\# \text{ employees, other industries, same area+1})$ | 0.042 ^c (0.022) | -0.028 (0.108) | -0.067 (0.107) | -0.022 (0.150) | 0.055 ^b (0.023) | 0.073 (0.116) | 0.034 (0.115) | 0.076 (0.162) |
| $\Delta \ln(\text{competition})$ | | | | | 0.002 (0.005) | 0.041 (0.030) | 0.048 (0.030) | 0.035 (0.043) |
| $\Delta \ln(\text{sectoral diversity})$ | | | | | 0.023 ^c (0.012) | -0.122 ^c (0.071) | -0.110 (0.071) | -0.125 (0.103) |
| Sargan-Hansen test/p-value | | | | 0.369 | | | | 0.414 |
| N | 61332 | 61332 | 61332 | 61332 | 61332 | 61332 | 61332 | 61332 |
| Centered R ² | 0.125 | 0.031 | 0.026 | 0.027 | 0.125 | 0.015 | 0.011 | 0.011 |

Note: Standard errors in parentheses. ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels.

(1) and (5) simple OLS, (2) and (6) are IV, with standard errors taking into account individual auto-correlation, (3) and (7) are GMM, with standard errors taking into account individual auto-correlation, (4) and (8) are GMM with Moulton standard errors.

Table 14: Instrumental variables approach, Naf 2-digit/Employment Area

| Model : | Dependent Variable: $\Delta \ln(\text{value added})$ | | | | | | | |
|---|--|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\Delta \ln(\text{employees})$ | 0.513 ^a (0.008) | 1.008 ^a (0.072) | 1.020 ^a (0.072) | 1.016 ^a (0.095) | 0.513 ^a (0.008) | 0.994 ^a (0.077) | 1.006 ^a (0.077) | 1.003 ^a (0.101) |
| $\Delta \ln(\text{capital})$ | 0.070 ^a (0.006) | 0.214 ^a (0.023) | 0.213 ^a (0.023) | 0.215 ^a (0.030) | 0.070 ^a (0.006) | 0.215 ^a (0.024) | 0.215 ^a (0.024) | 0.218 ^a (0.031) |
| $\Delta \ln(\# \text{ employees, other firms, same industry-area+1})$ | 0.008 ^a (0.002) | 0.054 ^c (0.032) | 0.057 ^c (0.031) | 0.050 (0.041) | 0.007 ^a (0.003) | 0.058 ^c (0.032) | 0.063 ^b (0.032) | 0.055 (0.041) |
| $\Delta \ln(\# \text{ employees, other industries, same area+1})$ | 0.002 (0.013) | -0.049 (0.113) | -0.048 (0.112) | -0.033 (0.159) | 0.000 (0.013) | -0.072 (0.138) | -0.080 (0.138) | -0.069 (0.191) |
| $\Delta \ln(\text{competition})$ | | | | | 0.006 (0.005) | 0.028 (0.038) | 0.032 (0.038) | 0.038 (0.054) |
| $\Delta \ln(\text{sectoral diversity})$ | | | | | -0.006 (0.011) | 0.067 (0.092) | 0.071 (0.092) | 0.085 (0.128) |
| Sargan-Hansen test/p-value | | | | 0.434 | | | | 0.346 |
| N | 62305 | 62305 | 62305 | 62305 | 62305 | 62305 | 62305 | 62305 |
| Centered R ² | 0.122 | 0-015 | 0-02 | 0-018 | 0.122 | 0-012 | 0-019 | 0-018 |

Note: Standard errors in parentheses. ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels.

(1) and (5) simple OLS, (2) and (6) are IV, with standard errors taking into account individual auto-correlation, (3) and (7) are GMM, with standard errors taking into account individual auto-correlation, (4) and (8) are GMM with Moulton standard errors.

Table 15: Instrumental variables approach, Naf 2-digit/Département

| Model : | Dependent Variable: $\Delta \ln(\text{value added})$ | | | | | | | |
|---|--|-------------------------------|--------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\Delta \ln(\text{employees})$ | 0.517 ^a (0.008) | 0.938 ^a (0.074) | 0.951 ^a (0.073) | 0.953 ^a (0.094) | 0.517 ^a (0.008) | 0.963 ^a (0.081) | 0.978 ^a (0.081) | 0.980 ^a (0.105) |
| $\Delta \ln(\text{capital})$ | 0.071 ^a (0.006) | 0.208 ^a (0.021) | 0.209 ^a (0.021) | 0.213 ^a (0.027) | 0.071 ^a (0.006) | 0.213 ^a (0.021) | 0.213 ^a (0.021) | 0.216 ^a (0.028) |
| $\Delta \ln(\# \text{ employees, other firms, same industry-area+1})$ | 0.016 ^a (0.004) | 0.150 ^a (0.054) | 0.155 ^a (0.052) | 0.151 ^b (0.072) | 0.016 ^a (0.004) | 0.184 ^a (0.054) | 0.186 ^a (0.053) | 0.184 ^b (0.075) |
| $\Delta \ln(\# \text{ employees, other industries, same area+1})$ | 0.021 (0.021) | -0.171 (0.112) | -0.196 ^c (0.111) | -0.164 (0.158) | 0.020 (0.022) | -0.174 (0.120) | -0.20 ^c (0.119) | -0.148 (0.175) |
| $\Delta \ln(\text{competition})$ | | | | | -0.002 (0.005) | -0.003 (0.035) | -0.002 (0.034) | -0.008 (0.052) |
| $\Delta \ln(\text{sectoral diversity})$ | | | | | -0.004 (0.017) | -0.071 (0.111) | -0.064 (0.110) | -0.10 (0.180) |
| Sargan-Hansen test/p-value | | | | 0.187 | | | | 0.153 |
| N | 64714 | 64714 | 64714 | 64714 | 64714 | 64714 | 64714 | 64714 |
| Centered R ² | 0.124 | 0.01 | 0.004 | 0.002 | 0.124 | 0.009 | 0.015 | 0.016 |

Note: Standard errors in parentheses. ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels.

(1) and (5) simple OLS, (2) and (6) are IV, with standard errors taking into account individual auto-correlation, (3) and (7) are GMM, with standard errors taking into account individual auto-correlation, (4) and (8) are GMM with Moulton standard errors.

Table 16: Agglomeration externalities and distance/Contiguity

| Model: | Dependent Variable: $\Delta \ln(\text{value added})$ | | |
|---|--|-------------------------------|--|
| | Employment area/Naf 3 | Département/Naf 3 | Département/Naf 2 |
| $\Delta \ln(\text{employees})$ | 0.869 ^a (0.109) | 0.946 ^a (0.097) | 0.960 ^a (0.106) |
| $\Delta \ln(\text{capital})$ | 0.216 ^a (0.031) | 0.188 ^a (0.031) | 0.232 ^a (0.032) |
| $\Delta \ln(\# \text{ employees, other firms, same industry-area}+1)$ | 0.038 (0.025) | 0.066 ^c (0.036) | 0.040 (0.039) |
| $\Delta \ln(\# \text{ employees, same industry, contiguous areas}+1)$ | 0.060 (0.052) | -0.028 (0.128) | 0.143 (0.114) |
| $\Delta \ln(\# \text{ employees, other industries, same area}+1)$ | 0.064 (0.206) | 0.142 (0.247) | -0.173 (0.178) |
| $\Delta \ln(\text{competition})$ | 0.047 (0.048) | 0.037 (0.045) | 0.035 (0.050) |
| $\Delta \ln(\text{sectoral diversity})$ | -0.079 (0.130) | -0.144 (0.113) | 0.121 (0.121) |
| Sargan-Hansen test/p-value | 0.846 | 0.414 | 0.523 |
| N | 54991 | 61332 | 62305 |
| Centered R ² | 0.021 | 0.012 | 0.015 |
| | | | 0.975 ^a (0.104) 0.217 ^a (0.027) 0.119 (0.089) 0.104 (0.113) -0.146 (0.149) -0.021 (0.052) -0.106 (0.171) 0.146 64714 0.004 |

Note: standard-errors in parentheses. ^a, ^b and ^c respectively denoting significance at the 1%, 5% and 10% levels. Standard-errors are Moulton's standard-errors.