Agglomeration Economies and Firm Level Labor Misallocation*

Lionel Fontagné † Gianluca Santoni ‡

November 30, 2016

Pre-print version of the paper forthcoming in the Journal of Economic Geography

Abstract

A large portion of the productivity differentials among locations is related to density. Firms located in denser areas are more productive due to agglomeration economies (Combes et al., 2012a). We provide in this paper an explanation of such economies: lower levels of input misallocation. The distribution of resources among heterogeneous firms has consequences for allocative efficiency, and denser areas provide a more favorable environment for dynamic matching between employers and employees. Using a methodology proposed by Petrin and Sivadasan (2013) we assess the degree of resource misallocation among firms within sectors for each of the 96 French "Départements" and 347 Employment Areas (*commuting zones*). Based on firm-level productivity estimates, we identify in the gap between the values of the marginal product and marginal input price the degree of input allocation at the firm level. Over the whole period 1993-2007 the average gap at firm level is around 9 thousands euro, with this gap increasing starting from the early 2000s. Importantly, firm misallocation is lower in denser areas, suggesting that the matching mechanism plays a part in explaining the productivity premium of agglomerated locations.

JEL classification: D24, R12, L25, O47

Keywords: Misallocation, Agglomeration, Productivity, Firm Level Data

*This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 320197 (Mapcompete project). The views expressed in this publication are the authors' alone, and do not necessarily reflect the views of the European Commission. We acknowledge support from the Bank of France at earlier stages of this project. We are grateful to Claire Lelarge and Thierry Mayer for helpful suggestions. We are also indebted to Irene Iodice for excellent research assistance.

†PSE - Université Paris 1 & CEPII. Correspondence: Centre d'Economie de la Sorbonne, 106-112 Bd de l'Hôpital, F-75647 Paris Cedex 13. Email: lionel.fontagne@univ-paris1.fr.

‡CEPII. Correspondence: CEPII, 113 rue de Grenelle 75007 Paris. Email: gianluca.santoni@cepii.fr.

Introduction

Denser areas are more productive. This might be due to selection since only the most productive firms locate in more competitive environments. It might be due also to agglomeration economies associated with better access to a variety of inputs, or the circulation of ideas. Combes et al. (2012a) show that firms located in denser areas are 9.7% more productive on average with respect to firms located in less dense environments. Their findings suggest that the main driver of this is not selection (i.e. tougher competition inducing less productive firms to exit the market) but agglomeration economies, based on the higher availability of services, infrastructure, and public goods (sharing), a denser labor market (matching), and technology spillovers (learning).[1]

A key driver of productivity is ease of resource allocation. The empirical literature confirms this critical property: resources (production inputs) do not flow freely from less to more productive firms, although more efficient firms are the most likely to survive in the market. Generally, firm level reallocation of economic activity tends to benefit highly productive (low cost) producers, resulting in an aggregate improvement although there are several factors that can hamper this continuous flow of resources from less to more efficient firms. These include business cycle[2], labor and capital rigidity, and the regulatory and competitive environments. The resulting resource misallocation implies that more efficient firms tend to be smaller than the optimal size while less efficient firms tend to be bigger than their optimum production scale. The dispersion of revenue-based productivity (the product of physical productivity and the firm's output price) reveals the degree of resource misallocation so-defined (Hsieh and Klenow, 2009). The rationale is that, in the absence of distortions, revenue-based productivity should be the same for all firms in the same sector. Alternatively, the difference between the marginal product value of each factor and its cost to the firm (Petrin and Sivadasan, 2013) could be considered. This difference or *gap* measures the degree of resource misallocation among firms within sectors. It measures the extent to which firms are not fully optimizing production.

From this perspective it is more straightforward to examine the channels which make agglomerated economies more productive, rather than looking at cross-country differences in the efficiency of resource allocation. Firms in denser areas - notwithstanding the distortions that are present in the economy as a whole such as labor

[1]Following the classification in Duranton and Puga (2004). Large cities can also benefit from the sorting of talents (Behrens et al., 2014)

[2]Lazear and Spletzer (2012) show that labor reallocation seems to be more conspicuous during expansionary periods than during recessions.

market rigidities - may match with more productive and better paid workers (Combes et al., 2012b). However, in relation to the difference between the value of the wage and the marginal product, a better matching should reduce the gap between the two at firm level. Using administrative data for the universe of legal units operating in the French manufacturing sector over the period 1993-2007, we show that this mechanism is at work: resource misallocation among firms within sectors is lower in denser *Départements*, and denser commuting zones (*Zones d'emploi*).

Our findings contribute to the large literature on the sizeable and persistent heterogeneity (i.e. dispersion) of firm productivity, even when productivity is computed within narrowly defined sectors.[3] The large firm level variability is not confined to productivity; e.g., U.S. sales growth rates show a standard deviation of about 50% (Davis et al., 2007), which for one third of the firms translates into expected growth of more than 60% and for one third of firms translates into an expected decline of more than 40%. High variability in firm productivity, sales, and entry and exit rates suggests that the allocation of resources plays an important role: notwithstanding the more structural employment shifts, the capacity of churning to drive resources towards the most efficient firms is conducive to aggregate performance.

If we focus on within-country productivity differentials, our findings relate also to international comparisons. A large portion of cross-country productivity differentials are imputable to input misallocation: in the case of heterogeneous firms, the distribution of resources among them has significant consequences for both allocation efficiency and aggregate outcomes.[4] The usual approach to measuring the degree of efficiency in resource allocation across countries is based on the covariance between firm size and productivity. If resources were allocated purely randomly this covariance would be zero; conversely the higher the covariance the more efficiently resources are allocated across firms (Bartelsman et al., 2009). [5] Market rigidity, regulatory distortions, and other frictions may weaken the correlation with fundamentals. Similarly, the empirical evidence reported in CompNet (Berthou and Sandoz, 2014)[6] shows that over the period 2003-2007 the distribution of inputs across

[3]Syverson (2004) reports a total factor productivity (TFP) ratio for the U.S. of 1.92 among firms in the 90th and the 10th percentile of the industry distribution. Within a narrow defined sector, most productive firms are able to produce almost twice the output of less productive ones, with the same amount of inputs. The degree of misallocation is even higher in China and India, the gain in TFP from achieving the same allocative efficiency as the U.S. would be between 30% and 50% for China, and as much as 40%-60% for India, while the increase in output would be almost twice that. See Hsieh and Klenow (2009). U.S. productivity inevitably involves gaps and a degree of misallocation, the distribution is used as the control group.

[4]See Hsieh and Klenow (2009), Syverson (2014), Dhingra and Morrow (2014).

[5]The procedure adopted which is in line with Olley and Pakes (1996)), uses the covariance between firm size and productivity within sectors to assess the efficiency of input allocation. Note that this is the static version of allocative efficiency, in a cross-section framework; see Haltiwanger (2011) for a discussion of static and dynamic allocative measures.

[6]The Competitiveness Research Network - CompNet - is composed of economists from the 28 central banks in the European Union (EU) plus the European Central Bank; international organizations (World Bank, Organisation for Economic Cooperation

European countries could improve significantly.[7]. More generally, Bento and Restuccia (2014) establish a clear relationship between observed international differences in the levels of resource misallocation and establishment size. Policy distortions, institutions, and market frictions are shown to be driving the extent of the misallocation.

A third strand of literature focuses on the dynamics of allocative inefficiency. Ranasinghe (2014) systematizes the idea proposed by Bento and Restuccia (2014) that one of the mechanisms at play is the impact of distortions on the incentive to invest to enhance productivity beyond the reallocation of resources within the firm. Firm productivity is endogenous and driven by investment decisions , conditional on the institutional environment of the firm. The reason for this is that policies affect heterogeneous firms differently, and shape their incentives to invest in future productivity differently.

Finally, the effect of the misallocation of resources which we examine through the lens of optimization of demand for labor at the establishment level extends beyond labor. Gopinath et al. (2015) study the misallocation of capital among firms in Spain in relation to financial frictions, and argue that it led to low productivity gains before the crisis. David et al. (2014) in the case of China and India, show how information frictions lead to capital misallocation. Restuccia and Santaeulalia-Llopis (2015) examine the impact of land misallocation in Malawi and the related small size of farms, on agricultural productivity. Efficient allocation would lead to a four-fold increase in aggregate productivity. Duranton et al. (2015) show how land misallocation in India translates into loan misallocation since land is used as collateral. They provide evidence of a cascade effect of misallocation of certain production factors within economies.

The rest of the paper is organized as follows. Section 1 presents the data and descriptive evidence of differences in firm productivity. The methodology described in Section 2 was inspired by Petrin and Sivadasan (2013). TFP estimation strategy is described in Section 2.1. Section 3 computes the value of the labor gaps at both sector (Section 3.1) and firm 3.2 level. In Section 4, we assess the effect of agglomeration economies on the dynamics of labor gaps, controlling for firm characteristics and the potential endogeneity of density variables. The last section concludes.

and Development - OECD, EU Commission), universities and think-tanks, and non-European central banks (Argentina and Peru) and organizations (U.S. International Trade Commission). The objective of CompNet is to develop a more consistent analytical framework for assessing competitiveness, allowing for a better correspondence between determinants and outcomes.

[7] The covariance between labor productivity and firm size reaches 0.2 for Hungary and Spain, meaning that in those countries labor allocation is about 20% more efficient than the random allocation benchmark; a similar analysis for the U.S. shows a correlation of about 50%. However, the results in Bartelsman et al. (2013) show a higher covariance for European countries, ranging from 15%-38%, confirming the existence of a sizeable efficiency gap with respect to the U.S. benchmark. This finding was challenged by Bellone and Mallen-Pisano (2013) who found a much smaller difference in the degree of factor misallocation between the U.S. and France.

1 Data and Productivity Estimation

Our evaluation of input allocation is performed using firm level balance sheet data to retrieve total factor productivity (TFP) estimations, from which we derive the marginal contribution of production inputs. Then, using firm (or industry) specific input prices it is possible to derive a monetary value for the firm level allocation inefficiencies. We use balance sheet data comprising information on the location of the establishment considered which defines two location measures: the *Département* level (there are currently 96 *Départements* within metropolitan France) and at the commuting zone level (*Zone d'emploi*). We ignore the municipality level which for our purposes is meaningless. There are good reasons for preferring one or other type of measure. The *Département* is an administrative category for firm location and was introduced in 1789. However, the corresponding geographical category might not be applicable to the contemporary economy and transport infrastructure. In contrast, commuting zones were defined for statistical purposes and were revised in 2011 based on the 2006 census. There are over 340 commuting zones which were defined jointly by the National Institute of Statistics (INSEE) and the Ministry of Labor.[8]

The main source of firm level data is the French Bénéfice Réel Normal (BRN)[9] dataset available from the fiscal administration. It contains balance-sheet information collected from firms' tax forms combined with detailed information on firms' balance sheets, including total, domestic, and export sales, and value added as well as many cost items including the wage bill, materials expenditure, etc. and the sector and location in which the firm operates. The dataset covers the period 1993-2007 and offers a very detailed representation of the aggregate economy. The fact that the information comes from the tax authorities ensures overall very high data quality. After excluding implausible observations i.e. those reporting negative or zero values for our variables of interest, and cleaning the data of potential outliers [10], we have an unbalanced panel of more than 115,000 manufacturing firms (of which 97,600 are single plant firms). [11] Figure 1 shows that while most firms in the sample have only one production plant, there are about 20% of companies which are multi-plant firms

[8]The INSEE definition is: "An employment zone is a geographical area within which most of the labor force lives and works, and in which establishments can find the main part of the labor force necessary to occupy the offered jobs."

[9]BRN is the normal tax regime for French firms.

[10]As a further robustness check we excluded observations with capital intensity or value added per worker above/below the 99th/1st percentile of the industry by year distribution. In fact, extreme values can be caused by misreporting but can also be induced by specific capital management strategies, e.g. an entrepreneur may create a separate entity which owns real estate assets, resulting in a large capital stock with few workers (see INSEE 2015, *Les entreprises en France*). See the results in table 14 in Appendix 6.3. Our main findings remain unaffected by changes in the capital intensity or value added per worker thresholds, and the exclusion of firms with fewer than 10 workers.

[11]We limit the analysis to the manufacturing sector to ease interpretation of the TFP estimation coefficients as marginal products; however, the underlying methodology can be applied also to other industries.

and this proportion is growing over time. It is interesting that 50% of the multi-plant firms are located in the same department (see Table 9 in the Appendix). The composition of our sample is shown in figure 1. It shows a stable proportion of exporters over time, suggesting that the empirical evidence discussed below is not driven by sample composition effects. The empirical international economics literature that exporters are significantly different from non-exporters on many dimensions (see Bernard et al. (2007) and Wagner (2012) for a recent survey).

[Figure 1 about here.]

Before turning to more sophisticated analysis we investigate whether in denser (more agglomerated) areas are more productive overall. The literature provides numerous examples of this pattern (see Combes et al. (2012a) for France). We are interested in confirming that our data exhibit this premium. We compute the TFP of single-plant firms, by sector (see below for a discussion of the method). Figure 2 plots the density of firms' TFP for two firm categories in 2000 in *Départements* below and above the median urbanization (see definition below). Our choice of the year 2000 is because it is in the middle of our time window; for obvious reasons, we are not interested in pooling years. However, note that to avoid simultaneity bias, the measure of urbanization is based on 1999. Figure 2 depicts the productivity premium for firms in denser *Départements*. This premium is 4.4%, significant at the 1% level. Although we do not include any controls, the difference is clear. We expect a more appropriate zoning of the French space to show a more marked difference, and recompute the same distributions at the level of commuting zones. The results in figure 3 confirm our conjecture; the premium is now 6.5% significant at the 1% level. This first descriptive evidence suggests that we should prioritize commuting zones; however, in what follows we retain both definitions.

[Figure 2 about here.]

[Figure 3 about here.]

2 Evaluating Misallocations at Firm Level

According to a well established line of reasoning, the distribution of resources among producers has a significant impact on a country's aggregate productivity and per capita income. As already noted, imperfections (or distortive regulation) in the input market can create incentives for less productive firms to produce beyond their optimal size while hindering the growth of the most efficient firms. The main effect is that the economy produces less than currently available resources will allow due only to their inefficient distribution. We extend this and argue that within an integrated economy with common institutions and regulation, which constitute a single market, allocative inefficiency among firms should be spread evenly within sectors across space. Observation of departures from this benchmark are indicative of geography and agglomeration effects within a country. Before discussing these effects, we present the method used to measure inefficiencies at the establishment level, keeping in mind that establishments are located at one point in space.

The following empirical work relies on Petrin and Sivasadan's methodology, used originally to evaluate the impact of a change in labor market regulation in Chile (Petrin and Sivadasan, 2013). Their approach, based on plant-level productivity estimates, aims to define the output loss due to inefficiencies in the allocation of inputs, and the impact of policy changes at both the firm and aggregate levels. The concept of firm specific "mis-allocation" refers to the gap between the value of the marginal product and the marginal input price. This gap is computed at firm level using the estimated coefficients from the TFP analysis, and can be further aggregated at the sector or spatial level. Moreover, since the is expressed in monetary terms, direct aggregation gives the amount of lost output due to the induced distortion in the distribution of resources across firms.

The underlying economic intuition is that in a context of perfect competition, the value of an input's marginal return should equate to its marginal cost. A wedge between the marginal return and the marginal cost is a sign that firms are not fully optimizing which limits aggregate output. Estimation of the gap with firm-level data employs a trans-logarithmic production function which allows us to control for the intensity of input use. The estimated production function for firm i at time t is:

$$q_{it} = \beta_l l_{it} + \beta_{ll} l_{it}^2 + \beta_k k_{it} + \beta_{kk} k_{it}^2 + \beta_{lk} l_{it} k_{it} + \omega_{it} + \varepsilon_{it} \tag{1}$$

Where q_{it} denotes value added, l_{it} is the number of employees and k_{it} is the fixed capital stock. All series are in logs and expressed in real terms. [12] As Loecker et al. (2012) note, the translog reduces to a Cobb-Douglas when we drop the squared and interaction terms, e.g. exclude the control for relative inputs intensity. Our results are robust to using a Cobb-Douglas production function.[13] However, in addition to its generality, the main advantage of the translog technology is that it produces elasticities that change by firm, year, and firm size. The elasticity of labor used to compute the misallocation index (the *gap*), is defined as: $\phi^l_{it} = \beta_l + 2\beta_{ll} l_{it} + \beta_{lk} k_{it}$.

The error term has two components: ω_{it} represents a Hicks-neutral productivity shock (observed by the firm but not by the econometrician), and ε_{it} it is uncorrelated with the input choice (unobservable to the firm and to the econometrician). The main complication here is that ω_{it} will affect the firm level input decision, inducing simultaneity bias to the production function estimation. The economic rationale relies on the fact that the present investments will be productive only in the next period, and a representative firm will choose how much to invest only after observing its current productivity level (see section 2.1 for a detailed discussion).

Estimation of equation (1) provides a measure of firm i's production efficiency - the difference between the observed and predicted level of output. Essentially, a firm is more productive with respect to another in the same sector if it can produce more output using the same level of inputs. [14] In order to set a benchmark output level we need to estimate the marginal return (i.e. marginal product) of each input for the representative firm from each industry. In what follows we focus on labor marginal productivity.[15] The marginal product of labor is given as the marginal increment in output per unit change in labor:

$$\frac{\partial Q_{it}}{\partial L} = \phi^l_{it} \frac{Q_{it}}{L_{it}} \tag{2}$$

Once the marginal product is recovered from the production function estimation – equation (2) – the value of the marginal product of labor is given by multiplying the marginal product by the firm level output price.

$$VMP^l_{it} = P_{kt} \left(\phi^l_{it} \frac{Q_{it}}{L_{it}} \right) \tag{3}$$

Since output prices at firm level are generally not available (or only for a sub-sample of the surveyed firms),

[12] Our empirical exercise uses industry price deflators from INSEE.
[13] See the results in table 12 in Appendix 6.2.
[14] Or reaching the same level of output using less input.
[15] It can be generalized to any input.

we use an industry specific price index, P_{kt}. Using industry prices introduces the risk of measurement error in prices (Foster et al., 2008). As noted by Petrin and Sivadasan (2013), marginal products of inputs, i.e. ϕ^z_{it}, are still consistent if the deviation of the plant level price from the industry price is not systematically correlated with the input levels; in this case our results should not be affected by omitted variables bias (i.e. unobserved prices) since we are using only within firm variation.

Finally, the degree of resource misallocation at firm level, the revenue to cost gap, is given by:

$$G^l_{it} = \mid VMP^l_{it} - w_{it} \mid \tag{4}$$

Where w_{it} represents the wage of the marginal worker in firm i. [16] To ease comparability over time, the value of G^l_{it} it is deflated using the consumer price index n(CPI)[17], the value of G^l_{it} in absolute terms expresses the increase in value added induced by an optimal reallocation of labor. In a setting where resources are allocated optimally and there are no frictions in the input markets, all firms will demand labor up to the point when the expected marginal return is equal to the marginal cost thus closing the gap. In reality there are several reasons why an economy might depart from such equilibrium: hiring and firing costs, capital adjustment costs, taxes, mark-ups, and management practices. According to Petrin and Sivadasan (2013) the social optimum is reached when all gaps are equal to zero, while an efficient allocation of labor implies that gaps are equated across firms (Syverson, 2011). Our goal is to check whether these gaps equalize across space within an integrated economy.

2.1 Productivity Estimation

In order to assess the input gaps the first step is to compute firm level TFP. Our measure of TFP is computed by employing Wooldridge (2009) implementation of the Levinsohn and Petrin (2003) algorithm using material inputs as a proxy for technology shocks, and considering labor as a freely adjustable (variable) input and capital as fixed. [18]

[16] Since we do not observe the salary paid to the marginal employee, we use the average wage as a proxy. Wages include salary and tax allowances.

[17] Results are robust to the use of the GDP deflator instead of CPI.

[18] The semi-parametric estimator proposed by Levinsohn and Petrin (2003) extends the methodology in Olley and Pakes (1996), which suggests the use of investments as a proxy to avoid the problem of simultaneity between a technology shock ω_{it} and the input choice in a two stage estimation procedure. [19] Levinsohn and Petrin suggest using raw materials as a proxy variable for ω_{it} mainly because investments are a valid proxy only if they adjust smoothly to productivity shocks, and also because intermediate goods tend to be reported more frequently in firm balance sheets. See Van Biesebroeck (2007) for a detailed discussion of the different methodologies used to estimate productivity (underlying assumptions and drawbacks).

Wooldridge suggests implementing Levinsohn and Petrin's methodology in a general method of moments (GMM) framework which takes account of potential contemporaneous error correlation among the two stages, in addition to heteroskedasticity and serial correlation. It is also robust to Ackerberg et al. (2006) critique[20]. It is worth noting that our findings are robust to different estimation methodologies, namely: the semi-parametric two stage estimator as well as the GMM implementation with labor as a fixed input. The latter case is particularly relevant since it assumes explicitly the existence of labor market frictions [21].

[Figure 4 about here.]

Figure 4 shows the distribution of firm-level TFP over selected years - 1995, 2000, and 2005. Over a decade, our estimations show that the productivity of French firms has increased significantly. In 2005 the average manufacturing firm was about 16.9% more productive than its 1995 counterpart (the difference in mean of the two distributions is statistically significant at the 1% level). The right shift in the distribution during the years suggests a not negligible redistribution of firms towards higher levels of productivity. However, this does not mean that the use of resources is increasingly close to optimal efficiency. Notwithstanding the developments in productivity, inefficiencies in factor allocation might hinder full enjoyment of the gains associated with technical progress. The within-industry productivity dispersion reveals a more heterogeneous picture; in 1995 the 90th to 10th percentile ratio for French manufacturing firms was 0.88, meaning that for a given amount of inputs, the most efficient firms were able to reach a level of production 142% higher than the least efficient firms[22]. In 2007, given the average increase in productivity, the interquartile ratios increased to 0.98 (i.e. 168%), suggesting that aggregate improvements were not driven by reallocation. Note that dispersion based on revenue productivity is usually smaller than if computed on quantity-based productivity (see Foster et al. (2008)); the reported values are likely to represent the lower bound of true sectoral variability.

3 Labor misallocation: aggregate and firm perspective

We next implement the method described in the previous section and present the evolution of labor gaps at the aggregate and firm levels.

[20]The main argument is that the coefficient of labor (or any other variable input) will not be identified in the two step Levinsohn and Petrin approach if its choice is a function of unobserved productivity.

[21]The results are available upon request.

[22] Since productivity is measured in log scale the percentage increase is given by $exp(0.85) - 1 = 142$.

3.1 Sectoral gaps

The marginal productivity of production inputs is reported in the first two columns in table 1. At the sectoral level, input elasticity is always positive and is estimated very precisely. Labor represents the highest coefficient in all industries as the input cost share. Estimated returns to scale are generally slightly below unity (on average 0.91); decreasing returns are a sufficient condition for an optimal input choice without adjustment. Having estimated the marginal productivity coefficients, computation of the resource allocation gap is straightforward - from equations (3) and (4). The main results for the labor return to cost wedge are reported in table 1.

[Table 1 about here.]

For a given sector, s at time t the mean absolute labor gap is defined as follows $Gap_s^{Abs} = \frac{\sum_{i \in s} |G_i^l|}{N_s}$. It measures the distance from the social optimum allocation[23] where each firm is operating under perfect competition, i.e. marginal revenue equal to marginal costs and no frictions in the input markets. For the whole manufacturing sector over the period 1994-2007 this figure is slightly above 8,700 euro per firm[24], with dispersion relatively high both between and also within industries, as shown by the coefficient of variation (CV). Instead of using perfect competition (zero gap) as the benchmark we are interested in knowing what the contribution to overall gains would be from achieving efficient allocation, i.e. the existing gaps are equal across all firms in a given sector. This results from the loosening of market constraints which allows the reallocation of one unit of labor (i.e. the marginal worker) across firms without changing the employment level and the structural frictions. This information is captured by the term shineff$_{\%}^{Abs}$, derived as the ratio: $1 - \frac{\sum_{i \in s} G_i^l / N_s}{Gap_s^{Abs}}$. For instance, in the case of machinery and equipment, resource allocation inefficiencies determine 50.2% of the mean absolute gap while structural frictions account for 49.8% of the overall gap. If we look at the last year in the sample we observe that about a fifth of the manufacturing firms in the sample report a positive wedge between the marginal return and cost of labor (see table 2).[25]

[Table 2 about here.]

The sign of the gap is meaningful since it helps to differentiate the variability from the direction of firm level misallocation. The average positive gap is roughly 34% higher than the negative counterpart, and the overall

[23] Under the implicit assumption that the marginal worker's productivity is in line with the average productivity in the observed firm.

[24] All monetary values are expressed in real terms (2005 euros), deflated using the CPI.

[25] The share of firms with positive gaps over the whole period is 22.2%.

distribution for positive wedges seems to be relatively more right skewed with respect to the negative wedges. Assuming an average labor cost[26] in France of 50,000 euro per year in 2007, an average negative wedge of 9,700 euro implies that the marginal return from labor is smaller than the cost of around 2.2 months of salary. Notice that there is a relatively high dispersion in the data since the coefficient of variation (cv) is 1.04. On the other hand, the value produced by the marginal worker is higher by almost 12,000 euro than its cost when a positive wedge is observed (with a cv of 1.32), although under perfect competition the firm should demand labor until its marginal return equals its cost.

3.2 Firm Level Evidence

In what follows we estimate the dynamics of the labor gap controlling for firm characteristics. We have two objectives. First we are interested in whether firms of different sizes face different obstacles in trying to optimize their use of labor. If the external labor market is sticky, firms may resort to internal markets, and especially in the case of large firms that rely on a large pool of internal competencies. Next, we investigate whether firms in denser areas exhibit lower labor gaps, controlling for firm productivity, export status, and multi-plant firms. The combination of the latter two sets of results shows clearly that, controlling for firm characteristics, denser areas provide better opportunities for matching employers and employees.

The baseline estimated equation is defined as:

$$Y_{it} = \alpha_0 + \delta_1 + \delta_2 + \delta_3 + \Gamma_{it}\beta + \xi_i + \upsilon_{it} \tag{5}$$

where Y_{it} is the value of the absolute labor gap, $\left|G^l_{it}\right|$. The time evolution of the dependent variable is accounted for by three sub-period dummies: δ_1 for the years 1998-2000, δ_2 for 2001-2003 and δ_3 for 2004-2007. The constant α_0 ccaptures the reference period gap value. The vector Γ_{it} includes a set of form and industry controls: log of firm age (linear and squared), a series of dummy variables identifying the quintile of firm turnover[27], a dummy for the export status (Exp_{it}) and an index for the degree of industry competition ($Comp_{st}$)[28]. Finally, ξ_i are firm fixed effects to control for unobserved heterogeneity and υ_{it} is an idiosyncratic

[26] Including both salary and tax allowances.

[27] The inclusion of a turnover quintile dummy is meant to control for firm productivity. The reference distribution is computed by sector and year.

[28] Sectoral competition is measured as the $ln(1/HH)_{st}$, where HH is an Herfindahl-Hirschmann index of sales concentration by sector s and year t

shock.

A potential source of concern is that we do not observe firm level prices and are obliged to rely on industry deflators. As noted by Gopinath et al. (2015), in fact, in the presence of markups the production function estimates may be biased downward and display decreasing returns. De Loecker (2011) posits that such bias emerges if the difference between the firm and the industry price is correlated systematically with the input choice[29], which requires proper controlling in the empirical analysis for the firm's pricing strategy. Bellone et al. (2016) shows both theoretically and empirically, that markups are positively correlated with firm productivity and export participation and negatively related to the toughness of competition. The set of controls included in the vector Γ_{it} controls directly for unobserved firm prices and variation in demand.[30]

The main results of our analysis on the evolution of the labor gap for manufacturing firms are reported in table 3. Column 1 shows the evolution of the gap conditional only on firm age and fixed effects; the estimated coefficients of the sub-period dummy show that the average wedge between the marginal return from and marginal cost of labor has increased significantly over time, especially in the last years of the sample. In 2004-2007, the average gap was around 2,000 euro higher compared to the reference period (1994-1997). In column 2 we add the covariates specified above to proxy for firms' pricing strategy, and the estimation results confirm the time evolution of the gap.

In the first two columns of table 3 the standard errors are clustered at firm level in order to deal with serial correlation; however, since productivity estimation is performed at the sector level this might induce cross-sectional dependence of the errors. In column 3 we assume that observations are correlated both over time and within sectors for a given year. [31] The results confirm that the average gap is increasing significantly over time, that more productive firms allocate labor more efficiently, and that conditional on the production quintile, exporting firms report a significantly smaller gap. [32]

The evolution over time of the labor gap for an average manufacturing firm with positive or negative values

[29]Since we use only within firm variation our baseline estimation should not be affected by the omitted (firm) price bias so long as firm to industry relative prices do not change over time.

[30]Loecker and Warzynski (2012) show how to retrieve firm level markups from production data, and propose that the output elasticity of a given input is equal to the expenditure share times the markup (i.e. markups are expressed as the difference between price and marginal cost). Unfortunately, since our dependent variable is built on a similar cost minimizing condition (the gap is measured as the difference between marginal revenue and marginal cost), including De Loecker and Warzynski measure of markups in the empirical specification introduces endogeneity bias.

[31]Following Cameron and Miller (2015) we use a two-way cluster-robust covariance matrix to correct standard errors.

[32]Note that the results do not change if we apply a 1 year lag to the covariates. Also, since it is beyond the scope of this paper we leave analysis of the relation between resource allocation and export/import participation to future research.

is depicted in figure 5[33]. The dynamics of the negative and positive gaps are different. A sharp increase in the average negative gap is observed from 2001 on. This has a large impact on the average absolute gap given the high frequency of negative gaps in the sample. This change is contemporary with the new labor market regulations but we cannot assess the causality. In contrast, the positive gap increased from the mid 1990s despite a short period of stabilization during the time considered.

In addition to identifying these two different evolutions, the method we use allows us to disentangle two possible categories of determinants of the observed gaps. For negative values, the lack of optimization could be driven by distortions hampering the firm's adjustment to the new market conditions. There are different explanations for the positive gaps which constrain firms to below their optimal size - as under perfect competition - due to market imperfections such as market power. Our results are robust to the labor gaps being restricted to negative values. Thus, the potential drawback related to our assumptions about competition are not driving our conclusions.

[Table 3 about here.]

[33]The two graphs report the value of the time dummies (interacted with an indicator variable for the characteristic of interest) holding all covariates at their mean value - equation (5).

[Figure 5 about here.]

In Table 4 we perform a series of robustness checks on the sensitivity of our results to sample selection. The evolution over time of firm input misallocations is consistent if we restrict the sample to firms with at least 20 employees ("restricted sample") or to small firms with fewer than 20 workers ("small firms"). Also, the evidence does not change if we restrict the sample to single plant firms which suggests that our results are not driven by compositional effects or measurement errors induced by consolidated financial accounts (in the case of multi-plant firms).

In many countries labor regulation is more binding for bigger firms. In France this increasingly stringent regulation is particularly relevant for firms with more than 50 workers. Above this threshold firms must organize a works council, set up a committee for working conditions (health and safety), and appoint a union representative. [34] The main effect of this increasing regulation is an increase in labor costs which may induce resource misallocation (see Garicano et al. (2013)) and potentially could affect our results. Column 4 shows that firms which remain below the 50 worker threshold, other things being equal, show the same time pattern as the whole sample

[Table 4 about here.]

4 Agglomeration economies and Labor Misallocation

We now address our central argument and enrich the baseline specification in equation 5 to test for the effect of agglomeration economies on return-cost wedges. Comparing the empirical firm productivity distribution across high and low density locations Combes et al. (2012a) show that there is a substantial efficiency premium associated with city size, and that this is even higher for highly productive firms. Interestingly, this premium is unrelated to selection and is driven by agglomeration economies. Combes et al. (2012a) distinguish selection from agglomeration externalities using a novel quantile approach which allows close comparison of the productivity distributions. Intuitively, this methodology relates the quantile of (log) productivity distribution in large and small cities to three key parameters: truncation, relative shift, and dilation. A standard prediction of firm heterogeneity models is that low productive firms will not survive in larger markets due to the higher level of

[34] Above this threshold firms are expected also to establish a "social plan" for of more than 9 employees are layed off at the same time, to show that the firm owner has tried to find alternative employment for those being dismissed.

competition: then productivity distributions should display a left truncation in denser areas. However, Combes et al. (2012a) find no evidence of left truncation (selection), instead, denser areas' productivity distributions appear to be right shifted (average productivity premium) and dilated (more productive firms benefit more) with these last two characteristics the result of the already mentioned agglomeration externality mechanisms - sharing, learning, and matching.

In what follows we focus on the matching channel, and test whether in denser areas the thicker labor market also affects firm resource allocation efficiency, i.e. return to cost wedge. To control for intra and inter industry agglomeration externalities we add to the vector Γ_{it} a set of measures on the economic environment at the Dpartement level (NUTS3 administrative entities) or at the Employment Area level. In defining the indicators we follow Martin et al. (2011): for a firm i located in the *Département* d and operating in the sector s we include:

- $Urbanization = ln\left(employees^{d}_{t} - employees^{ds}_{t} + 1\right)$
- $Location = ln\left(employees^{ds}_{t} - employees^{ds}_{it} + 1\right)$

where *Urbanization* is the number of employees in other industries within the same *Département* d of industry i. This variable is meant to capture inter-industry externalities, measured as the size of other industries' employment. *Location*, refers to intra-industry externalities measuring the number of employees working in the same industry s and the same *Département* d as firm i.[35] In order to limit measurement errors stemming from the inclusion of multi-plant firms we restrict the sample to single-plant firms (roughly 80% of the sample). Indicators are computed similarly for the Employment Areas.

Baseline OLS results are reported in columns 1 and 2 of table 5 and suggest that on average, denser areas are associated with lower gaps while highly specialized Departments do not seem to show sizeable differences in the value of the labor gap. However, we think that our baseline estimation may suffer from endogeneity bias since the distribution of skills may differ across employment areas (or departments). Combes et al. (2012b) examine the distribution of skills across employment areas in France and find strong and significant evidence that average workers in denser areas are more skilled. Worker sorting by skills across cities could induce attenuation bias in the OLS estimation because the value of the (true) marginal product of labor is unobservable at firm level, and

[35] Martin et al. note that from a firm perspective the two measures, combined with the firm's actual number of employees, fully describe local manufacturing employment.

is based on the input elasticities estimated at sectoral level, whereas we are able to observe the average wage at firm level. Thus, workers in denser areas may be more productive (unobserved) and receive higher wages (observed); as a consequence we may overestimate the gap in denser areas inducing attenuation bias on the OLS coefficient of density.

In order to control for this source of bias we adopt the following strategy. We take first differences of our baseline equation to remove the additive firm specific fixed effect, and then estimate the transformed model using two year lagged values of $Urbanization_{t-2}$ and $Location_{t-2}$ as instruments for first-differenced variables. As suggested by Martin et al. (2011) in the case of a convergence process differenced variables should be negatively correlated to past levels making them relevant instruments. We also include the lagged number of establishments in the same *Département*, sector and year as additional instrument to allow us to perform a Hansen test on the assumption that past shocks are orthogonal Δv_{it}.

[Table 5 about here.]

Instrumental variables results are reported in columns 3 and 4 of table 5. The estimated coefficients of urbanization are larger in magnitude and statistically significant, confirming our previous findings for the effect of agglomeration economies on the allocation of labor at firm level. In columns 2 and 4 we allow for cross-sectional dependence of the error terms by location-year, and serial correlation.[36] In terms of the magnitude of the effect a 10% increase in the degree of urbanization is associated with a decrease in the average gap of roughly 243 euro[37]. The reported first stage F-test rejects a weak identification problem, whereas the Hansen pvalue supports the validity of our instruments, pointing to a causal interpretation of the effect of urbanization on labor allocation.

In order to test the robustness of our results to the level of geographical aggregation which has been shown potentially to affect estimation through the well know "Modifiable Area Unit Problem" (MAUP)[38], we run our analysis using "Employment Areas" (EA) as geographical unit of interest instead of *départments*.

We use the EA zoning established in year 1990 which organizes national territories into non overlapping commuting zones. The fact that EA boundaries have been defined prior to the estimation period should limit

[36]The analysis is robust to a more general clustering by sector and location, results are reported in table 13 in the Appendix 6.3.

[37]Note that the estimated model is level-log.

[38]Briant et al. (2010) show that the size of geographical units may affect the statistical results if the dependent variable is at a different level of aggregation. However, the results using French employment areas also indicate that this potential source of bias is second order with respect to model mis-specification.

possible simultaneity bias between the current distribution of economic activities and EA zoning. The results of the EA regressions are reported in table 6 both OLS and IV estimations confirm that our main findings are robust to changes in geographical units.

[Table 6 about here.]

5 Conclusion

Firms in denser areas are more productive. We argue that the gap between the value of the marginal product and marginal input price which reveals inefficiencies in input allocation is reduced in agglomerated locations. The nice feature of this approach using reasoning at the margin, is to give monetary value to this misallocation and to disentangle positive and negative gaps. Using a methodology proposed by Petrin and Sivadasan (2013) we were able to assess the degree of resource misallocation at the firm level using French administrative data. The location of the firm (within French Dpartements or within Employment Areas) is observed which provides information on the degree of misallocation within sectors among locations of different density conforming to a common regulatory framework (e.g. labor market regulations). The average (marginal) gap at firm level over the period 1993-2007 is close to 9,000 euro.

We confirmed that misallocation has a spatial dimension: resource allocation and the associated effect on productivity is related not only to firms' characteristics but also to the environment in which they operate. Denser locations offer a better match between employers and employees. Urbanization at *Départment* level and at the Employment Area as well, is associated with a lower average labor gap, suggesting that such matching is playing a role in determining the productivity advantage of denser areas.

Bibliography

Ackerberg, D., Caves, K., and Frazer, G. (2006). Structural identification of production functions. MPRA Paper 38349, University Library of Munich, Germany.

Bartelsman, E., Haltiwanger, J., and Scarpetta, S. (2013). Cross-country differences in productivity: The role of allocation and selection. *The American Economic Review*, 103(1):305–334.

Bartelsman, E. J., Haltiwanger, J. C., and Scarpetta, S. (2009). Cross-country differences in productivity: The role of allocation and selection. Working Paper 15490, National Bureau of Economic Research.

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3):507–553.

Bellone, F. and Mallen-Pisano, J. (2013). Is Misallocation Higher in France than in the United States? GREDEG Working Papers 2013-38, Groupe de REcherche en Droit, Economie, Gestion (GREDEG CNRS), University of Nice Sophia Antipolis.

Bellone, F., Musso, P., Nesta, L., and Warzynski, F. (2016). International trade and firm-level markups when location and quality matter. *Journal of Economic Geography*, 16(1):67–91.

Bento, P. and Restuccia, D. (2014). Misallocation, establishment size, and productivity. Working paper tecipa-517, University of Toronto, Department of Economics.

Bernard, A. B., Jensen, J. B., Redding, S. J., and Schott, P. K. (2007). Firms in international trade. *The Journal of Economic Perspectives*, 21(3):105–130.

Berthou, A. and Sandoz, C. (2014). Labour productivity in europe: allocative efficiency of labour or performance of firms? Bulletin Trimestriel 34, Banque de France.

Briant, A., Combes, P.-P., and Lafourcade, M. (2010). Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? *Journal of Urban Economics*, 67(3):287–302.

Cameron, A. C. and Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2):317–373.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., and Roux, S. (2012a). The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection. *Econometrica*, 80(6):2543–2594.

Combes, P.-P., Duranton, G., Gobillon, L., and Roux, S. (2012b). Sorting and local wage and skill distributions in france. *Regional Science and Urban Economics*, 42(6):913–930.

David, J. M., Hopenhayn, H. A., and Venkateswaran, V. (2014). Information, misallocation and aggregate productivity. Working Paper 20340, National Bureau of Economic Research.

Davis, S. J., Haltiwanger, J., Jarmin, R., and Miranda, J. (2007). Volatility and dispersion in business growth rates: Publicly traded versus privately held firms. In *NBER Macroeconomics Annual 2006, Volume 21*, pages 107–180. MIT Press.

De Loecker, J. (2011). Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity. *Econometrica*, 79(5):1407–1451.

Dhingra, S. and Morrow, J. (2014). Monopolistic competition and optimum product diversity under firm heterogeneity. mimeo.

Duranton, G., Ghani, S. E., Goswami, A. G., and Kerr, W. R. (2015). Effects of land misallocation on capital allocations in india. Technical Report 7451.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In Henderson, J. V. and Thisse, J. F., editors, *Handbook of Regional and Urban Economics*, volume 4 of *Handbook of Regional and Urban Economics*, chapter 48, pages 2063–2117. Elsevier.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review*, 98(1):394–425.

Garicano, L., LeLarge, C., and Reenen, J. V. (2013). Firm size distortions and the productivity distribution: Evidence from france. Working Paper 18841, National Bureau of Economic Research.

Gopinath, G., Kalemli-Ozcan, S., Karabarbounis, L., and Villegas-Sanchez, C. (2015). Capital Allocation and Productivity in South Europe. Working Papers 728, Federal Reserve Bank of Minneapolis.

Haltiwanger, J. (2011). *Innovation Policy and the Economy*, volume 12, chapter Job Creation and firm dynamics in the US, pages 17–38. University of Chicago Press.

Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, 124(4):1403–1448.

Lazear, E. P. and Spletzer, J. R. (2012). Hiring, churn and the business cycle. *American Economic Review: Papers and Proceedings*, 102(3):575–579.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies*, 70(2):317–341.

Loecker, J. D., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2012). Prices, Markups and Trade Reform. NBER Working Papers 17925, National Bureau of Economic Research, Inc.

Loecker, J. D. and Warzynski, F. (2012). Markups and Firm-Level Export Status. *American Economic Review*, 102(6):2437–71.

Martin, P., Mayer, T., and Mayneris, F. (2011). Spatial concentration and plant-level productivity in france. *Journal of Urban Economics*, 69(2):182–195.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Petrin, A. and Sivadasan, J. (2013). Estimating lost output from allocative inefficiency, with an application to chile and firing costs. *Review of Economics and Statistics*, 95(1):286–301.

Ranasinghe, A. (2014). Impact of policy distortions on firm-level innovation, productivity dynamics and tfp. *Journal of Economic Dynamics and Control*, 46:114–129.

Restuccia, D. and Santaeulalia-Llopis, R. (2015). Land Misallocation and Productivity. Working Papers tecipa-533, University of Toronto, Department of Economics.

Syverson, C. (2004). Product substitutability and productivity dispersion. *Review of Economics and Statistics*, 86(2):534–550.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature*, 49(2):326–65.

Syverson, C. (2014). The importance of measuring dispersion in firm-level outcomes. *IZA World of Labor*, 53(.):.

Van Biesebroeck, J. (2007). Robustness of productivity estimates*. *The Journal of Industrial Economics*, 55(3):529–569.

Wagner, J. (2012). International trade and firm performance: a survey of empirical studies since 2006. *Review of World Economics*, 148(2):235–267.

Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters*, 104(3):112–114.

6 Appendix

6.1 Data & First Stage Estimation

The dataset covers the period 1993-2007. The end point of 2007 was chosen to exclude the contrasting reactions of firms to the subsequent economic crisis. After excluding implausible observations, namely those reporting negative or zero values for our variables of interest, and cleaning the data of potential outliers, we have an un-balanced panel of 109,161 firms in the French manufacturing sector.[39]. Single plant firms represent 80% of yearly observations meaning that in the vast majority of the cases we observe production functions at plant level; however, the share of single plant firms is slightly decreasing over time, see 9.

Table 8 reports the descriptive statistics for the estimation sample, monetary values are expressed in thousand euro (deflated using industry level prices), the labor gap is expressed in real 2005 euro (deflated using the CPI).

[Table 7 about here.]

[Table 8 about here.]

[Table 9 about here.]

[Table 10 about here.]

[39] We limit the analysis to the manufacturing sector to ease interpretation of the TFP estimation coefficients as marginal products; however, the underlying methodology can be applied also to other industries.

6.2 Robustness I: Alternative total factor productivity estimations

In the following section we report the results obtained using different approaches to the computation of total factor productivity (TFP) and the implied marginal productivity of labor. Estimation of the firm specific labor gap starts with a Cobb- Douglas production function for firm i at time t defined as follow:

$$q_{it} = \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \varepsilon_{it} \tag{6}$$

where q_{it} denotes value added, l_{it} the number of employees, k_{it} the fixed capital stock and m_{it} is demand for intermediates. Consistent with the translog specification productivity is approximated by a function of the observable variables such as capital (k_{it}) and intermediates (m_{it}), and estimated using Wooldridge (2009) implementation of Levinsohn and Petrin's framework. Table 11 and 12 report the results for the baseline and the agglomeration economies using the labor gap computed from the input elasticities estimated from equation 6. In both cases the empirical evidence is confirmed suggesting that our main results are not driven by the functional form of the production function. In addition, Figure 6 corroborates the evidence found for the different pattern between positive and negative wedges.

[Table 11 about here.]

[Figure 6 about here.]

[Table 12 about here.]

6.3 Robustness II: Clustering and Sample Sensitivity

In this final section we test our inferences against the different levels of clustering in table 13 where we apply a more general clustering correction for the sector-geographic dimension which allows the errors term to be correlated arbitrarily across time, sector, and location. [40] The results confirms that our main variable of interest is always significant (although less precisely estimated in column 2) suggesting that our inference is robust to the assumption of a more general variance-covariance matrix.

Table 14 shows that our results are not driven by measurement errors in the firm level data or by very small firms. Columns 1 and 3 exclude all observations with a capital intensity or value added per worker above/below the 99th/1st percentile of the relative industry distribution; columns 2 and 4 further exclude firms with fewer than 10 workers. Again we find robust and consistent evidence suggesting that estimation of the gap (and the underlying production function parameters) seems not to be driven by outliers, very small firms, or the particular functional form we assume for the production function.

[Table 13 about here.]

[Table 14 about here.]

[40] In our preferred specification in the main text we cluster at firm and sector-geographic-year levels implying that the errors can be correlated across time for a given firm, and across firms (within the same sector and location) for a given year.

List of Figures

Figure 1: Distribution of firms in the estimation sample, number of exporters and single-plant firms (manufacturing only)

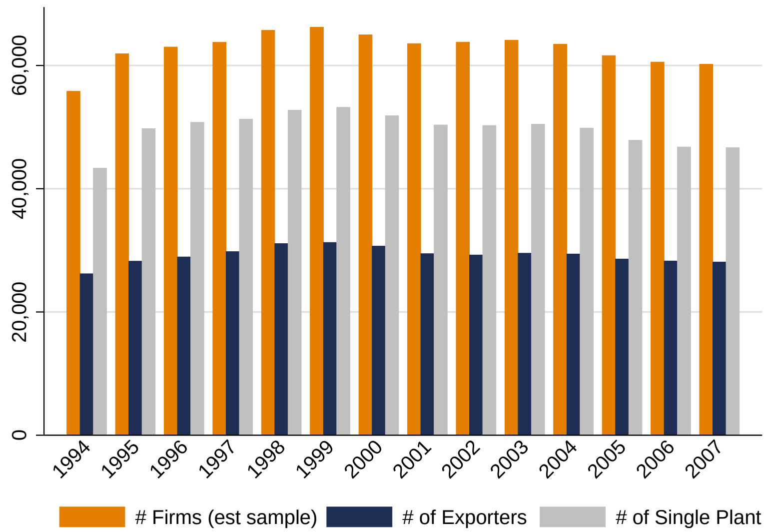


Figure 2: Urbanization and productivity by départements, single-plant firms (manufacturing only)

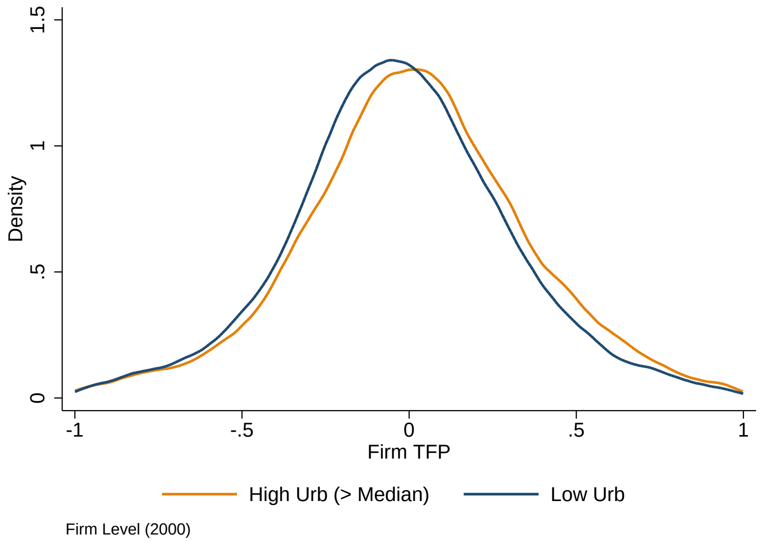


Figure 3: Urbanization and productivity by employment areas, single-plant firms (manufacturing only)

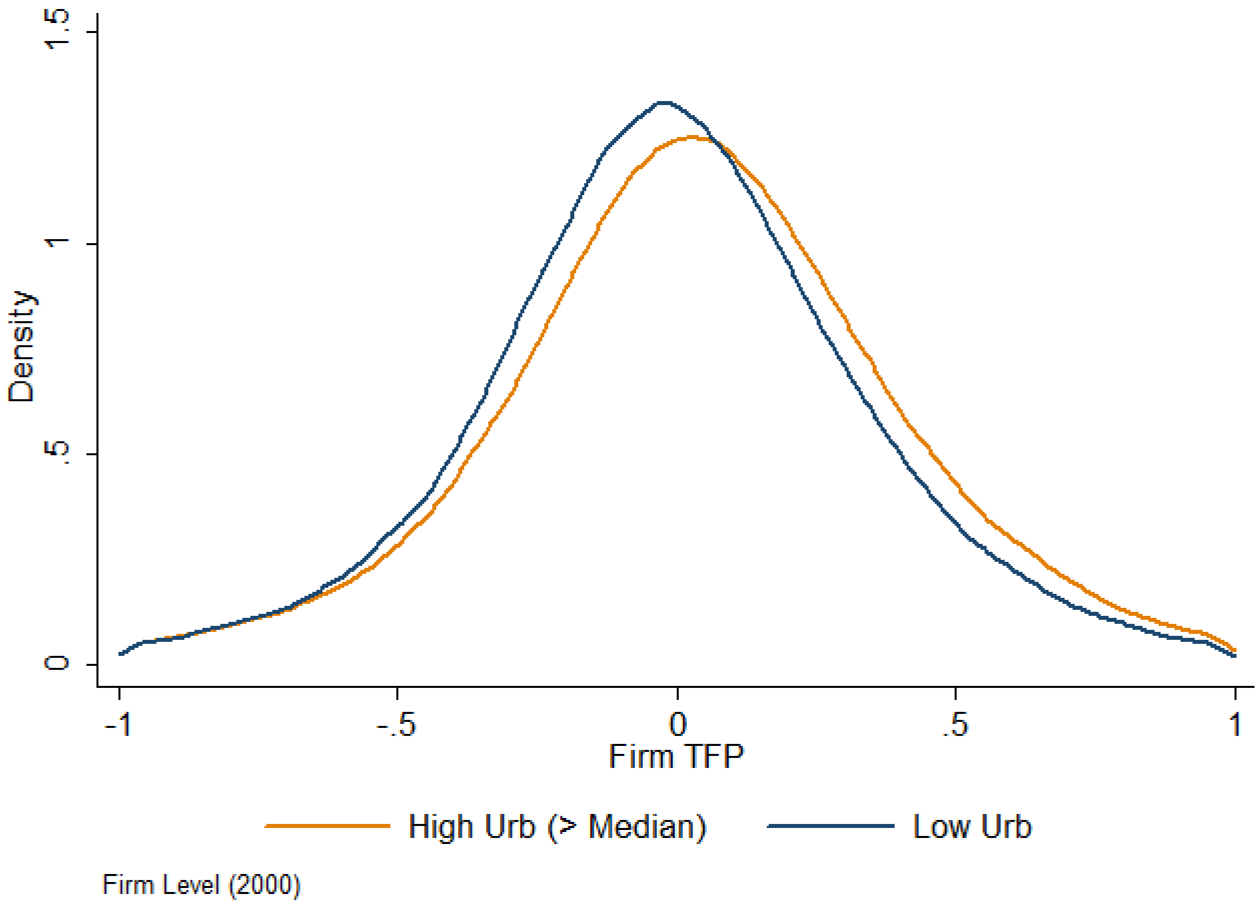
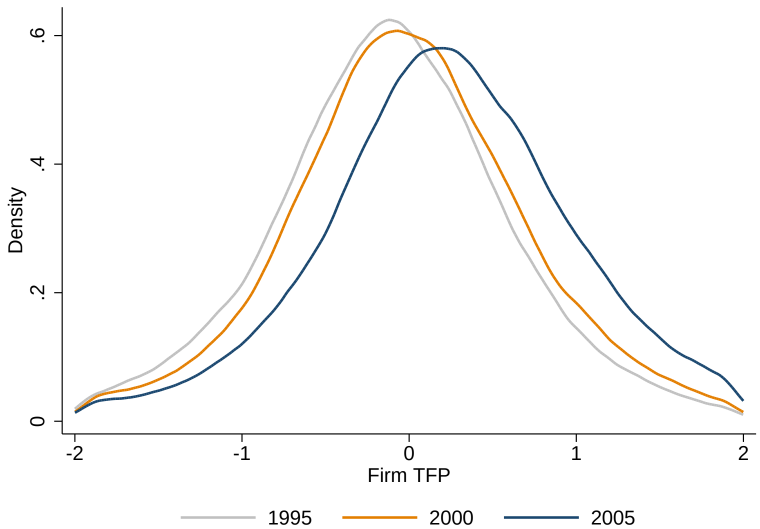


Figure 4: Manufacturing firms' productivity distribution , ω_{it}



Note: The graph reports the distribution of manufacturing firm TFP for selected years. Pooled distribution is standardized (over the whole period) to have zero mean and standard deviation equal to 1. The three distributions are statistically different at the 1% confidence level.

Figure 5: Average Labor Gap conditional on firm characteristics

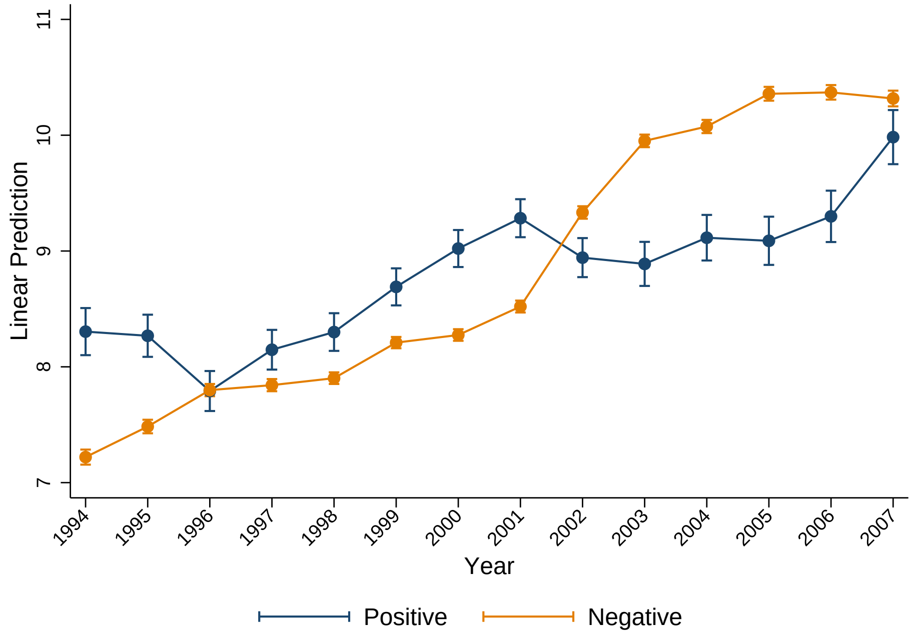
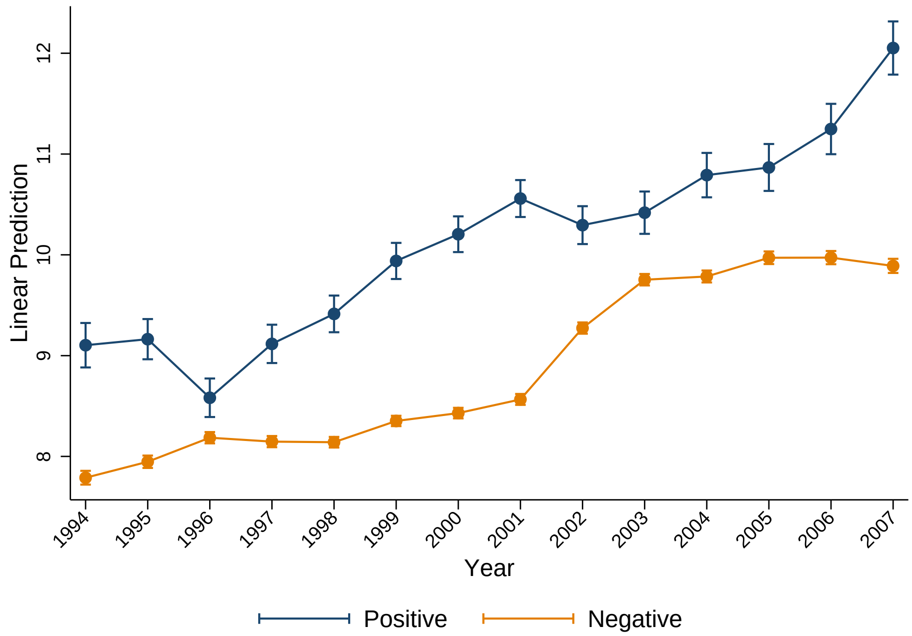


Figure 6: Average labor Gap conditional on firm characteristics



List of Tables

Table 1: Average Absolute Labor Gap by sector – years 1994-2007

Industry	Input Coefficients and RTS				Gap^{Abs}				Number
	ϕ^l	ϕ^k	RTS	Sd RTS	Mean	CV	$shineff^{Abs}_{\%}$	$Pos_{\%}$	Obs
Basic metals	0.629	0.223	0.851	0.058	9.052	0.994	38.3	19.4	8557
Beverages	0.645	0.349	0.994	0.042	18.488	1.095	33.8	66.0	9836
Chemicals	0.642	0.218	0.860	0.030	12.572	1.155	96.1	36.8	20788
Computer and Elect	0.558	0.164	0.722	0.036	14.123	0.776	23.6	12.2	25965
Electrical Equip	0.600	0.279	0.879	0.053	10.449	0.837	29.9	13.1	17229
Fabricated metal	0.658	0.269	0.927	0.019	8.226	0.894	38.1	17.8	145354
Food products	0.636	0.268	0.904	0.041	6.644	1.093	62.4	27.7	133132
Furniture	0.616	0.272	0.889	0.040	8.944	0.753	22.9	11.1	29987
Leather products	0.740	0.308	1.049	0.038	7.056	1.320	76.7	28.2	8937
Machinery and Equip	0.662	0.210	0.872	0.018	9.794	0.968	50.6	21.0	53899
Motor vehicles	0.658	0.201	0.859	0.067	8.727	1.088	52.2	19.9	14335
Non-metallic pro	0.604	0.297	0.901	0.030	9.525	0.998	57.8	20.5	33452
Other Manuf	0.648	0.292	0.941	0.038	9.897	0.970	59.9	23.8	37700
Other transport	0.683	0.221	0.904	0.046	8.656	1.115	56.9	25.8	6206
Paper products	0.641	0.274	0.915	0.019	8.747	1.028	59.4	25.0	15449
Pharmaceutical	0.475	0.257	0.732	0.051	19.022	0.897	63.4	21.9	4300
Printing and rec	0.671	0.234	0.905	0.024	9.142	0.954	36.4	17.3	66649
Repair and instal	0.698	0.174	0.873	0.024	8.031	1.025	49.1	20.8	74799
Rubber and plastic	0.625	0.251	0.877	0.040	8.306	1.003	56.6	23.7	38555
Textiles	0.650	0.335	0.985	0.061	8.815	1.140	60.9	22.8	26441
Wearing apparel	0.713	0.331	1.044	0.083	9.648	1.196	85.0	29.8	33714
Wood products	0.655	0.284	0.939	0.040	6.457	1.033	48.7	22.8	37475

Table 2: Labor Gap decomposition, year 2007

	$\lvert\tilde{G^l_{it}}\rvert$	$G^l_{it} > 0$	$G^l_{it} < 0$
# of Firms	55850	12700	43150
Share (%)	100	23	77
Mean	9.680	12.031	8.988
sd	10.065	15.836	7.429
10%	1.679	0.805	2.150
Median	7.211	6.036	7.394
90%	19.049	31.031	16.964

Table 3: Evolution of labor Gap by selected period, real euro (thousand)

Dep. Var. :	Labor Gap $\lvert G^{l}_{it}\rvert$		
	(1)	(2)	(3)
1998-2000	-0.048**	-0.040**	-0.040
	(0.020)	(0.020)	(0.092)
2001-2003	1.080***	1.050***	1.050***
	(0.027)	(0.027)	(0.091)
2004-2007	2.112***	2.039***	2.039***
	(0.038)	(0.038)	(0.086)
Size: 2^{nd} quintile		-1.421***	-1.421***
		(0.064)	(0.074)
Size: 3^{rd} quintile		-2.458***	-2.458***
		(0.087)	(0.113)
Size: 4^{th} quintile		-3.258***	-3.258***
		(0.114)	(0.153)
Size: 5^{th} quintile		-4.197***	-4.197***
		(0.153)	(0.211)
$Comp_{st}$		0.115***	0.115**
		(0.019)	(0.047)
Exp_{it}		-0.214***	-0.214***
		(0.032)	(0.034)
$ln(age)_{it}$	0.200*	0.659***	0.659***
	(0.120)	(0.120)	(0.160)
$ln(age)^2_{it}$	-0.033*	-0.106***	-0.106***
	(0.019)	(0.020)	(0.026)
Cluster Level	i	i	i & st
Observations	830,462	830,462	830,462
# of Firms	103,046	103,046	103,046
R-squared	0.599	0.601	0.601

Standard errors in parenthesis: *, **, *** denote statistical significance at the 10%, 5%, and 1% level, respectively. All regressions include firm fixed effects. Dependent variable: labor gap in real euro. Marginal productivity of labor is computed using Wooldridge (2009) modification of the Levinsohn-Petrin algorithm which considers capital as a fixed input, labor as a flexible input, raw materials as a proxy for unobserved productivity shocks, and a translog production function.

Table 4: Evolution of labor Gap sample sensitivity

Dep. Var :	Labor Gap $\lvert G^l_{it} \rvert$			
	Restricted sample	Small firms	Single Plant	$\leq$ 49 Workers
1998-2000	0.090	-0.138	-0.064	-0.084
	(0.100)	(0.086)	(0.091)	(0.091)
2001-2003	1.171***	0.901***	1.003***	1.011***
	(0.096)	(0.087)	(0.093)	(0.092)
2004-2007	2.313***	1.718***	1.956***	1.953***
	(0.095)	(0.100)	(0.090)	(0.086)
...				
Controls	Yes	Yes	Yes	Yes
Cluster Level	i & st	i & st	i & st	i & st
Observations	256,029	568,823	662,630	723,055
# of Firms	32,145	78,475	87,303	93,139
R-squared	0.677	0.596	0.596	0.594

Standard errors in parenthesis: *, **, *** denote statistical significance at the 10%, 5%, and 1% level, respectively. All regressions include firm fixed effects. Dependent variable: labor gap in real euro. Marginal productivity of labor is computed using Wooldridge (2009) modification of the Levinsohn-Petrin algorithm which considers capital as a fixed input, labor as a flexible input, raw materials as a proxy for unobserved productivity shocks, and a translog production function.

Table 5: Agglomeration Externality: *département*

Dep. Var.:	Labor Gap $\|G_{it}^{l}\|$			
	(1)	(2)	(3)	(4)
	OLS		GMM-IV	
$Urbanization_{dst}$	-0.280*	-0.280*	-2.437**	-2.418***
	(0.151)	(0.148)	(1.106)	(0.818)
$Location_{idst}$	0.010	0.010	-0.297	-0.254
	(0.051)	(0.050)	(0.444)	(0.331)
SIZE: 2^{nd} quintile	-1.330***	-1.330***	-1.272***	-1.269***
	(0.078)	(0.068)	(0.095)	(0.081)
SIZE: 3^{rd} quintile	-2.299***	-2.299***	-2.259***	-2.246***
	(0.117)	(0.096)	(0.142)	(0.114)
SIZE: 4^{th} quintile	-3.007***	-3.007***	-2.948***	-2.918***
	(0.164)	(0.128)	(0.197)	(0.144)
SIZE: 5^{th} quintile	-3.842***	-3.842***	-3.641***	-3.603***
	(0.225)	(0.174)	(0.267)	(0.183)
$Comp_{st}$	0.021	0.021	-0.018	-0.015
	(0.029)	(0.021)	(0.022)	(0.020)
Exp_{it}	-0.188***	-0.188***	-0.123***	-0.122***
	(0.034)	(0.034)	(0.036)	(0.034)
$ln(age)_{it}$	-0.074	-0.074	-0.972***	-0.981***
	(0.162)	(0.145)	(0.240)	(0.279)
$ln(age)_{it}^{2}$	-0.011	-0.011	0.171***	0.171***
	(0.026)	(0.024)	(0.045)	(0.051)
FEs	i & t	i & t	i & t	i& t
Cluster Level	i & st	i & dst	i & st	i & dst
Observations	662,630	662,630	460,320	460,320
# of Firms	87,303	87,303	74,646	74,646
First Stage F-test			43.38	59.19
Hansen J (p-value)			0.415	0.170

Standard errors in parenthesis: *, **, *** denote statistical significance at the 10%, 5%, and 1% level, respectively. All regressions include firm and year fixed effects. Dependent variable: labor gap in real euro. Single plant firms only. Results are robust using the degree of competition at the sector level, $Comp_{dst}$. First stage F test refers to the Kleibergen-Paap F statistics.

Table 6: Agglomeration externality: employment areas

Dep. Var.:	Labor Gap $\|G^l_{it}\|$			
	(1)	(2)	(3)	(4)
	OLS		GMM-IV	
$Urbanization_{dst}$	-0.140*	-0.140*	-2.490***	-2.515***
	(0.076)	(0.078)	(0.913)	(0.669)
$Location_{idst}$	0.017	0.017	-0.115	-0.110
	(0.027)	(0.027)	(0.146)	(0.135)
SIZE: 2^{nd} quintile	-1.332***	-1.332***	-1.271***	-1.271***
	(0.078)	(0.069)	(0.095)	(0.081)
SIZE: 3^{rd} quintile	-2.300***	-2.300***	-2.254***	-2.252***
	(0.117)	(0.095)	(0.143)	(0.113)
SIZE: 4^{th} quintile	-3.009***	-3.009***	-2.928***	-2.922***
	(0.164)	(0.127)	(0.198)	(0.142)
SIZE: 5^{th} quintile	-3.845***	-3.845***	-3.617***	-3.610***
	(0.225)	(0.172)	(0.268)	(0.180)
$Comp_{st}$	0.020	0.020	-0.015	-0.014
	(0.029)	(0.021)	(0.023)	(0.020)
Exp_{it}	-0.189***	-0.189***	-0.123***	-0.123***
	(0.034)	(0.034)	(0.036)	(0.033)
$ln(age)_{it}$	-0.075	-0.075	-0.945***	-0.945***
	(0.162)	(0.144)	(0.243)	(0.276)
$ln(age)^2_{it}$	-0.011	-0.011	0.166***	0.165***
	(0.026)	(0.023)	(0.045)	(0.050)
FEs	i & t	i & t	i & t	i& t
Cluster Level	i & st	i & dst	i & st	i & dst
Observations	662,630	662,630	460,320	460,320
# of Firms	87,303	87,303	74,646	74,646
First Stage F-test			28.41	36.67
Hansen J (p-value)			0.845	0.744

Standard errors in parenthesis: *, **, *** denote statistical significance at the 10%, 5%, and 1% level, respectively. All regressions include firm and year fixed effects. Dependent variable: labor gap in real euro. Single plant firms only. Results are robust using the degree of competition at the sector level, $Comp_{dst}$. First stage F-test refers to the Kleibergen-Paap F statistics.

Table 7: Number of firms in the estimation sample (by year)

Year	Firms	% Single Plant	% Within Same Dep.
1994	53442	0.791	0.938
1995	60426	0.804	0.943
1996	61205	0.809	0.945
1997	61672	0.809	0.946
1998	63020	0.808	0.947
1999	63346	0.810	0.948
2000	61979	0.805	0.947
2001	60271	0.801	0.947
2002	60154	0.798	0.946
2003	60236	0.799	0.947
2004	59372	0.798	0.947
2005	56863	0.790	0.946
2006	55819	0.789	0.945
2007	52657	0.787	0.945

Table 8: Descriptive statistics, firm level variables

variable	Obs	Mean	Sd	min	p50	max
Labor Gap $\|G_{it}^{l}\|$	830462	8.767	9.147	0.150	6.565	71.599
Age	830462	19	16	0	15	107
Export Status	830462	0.45	0.50	0	0	1
VA	830462	2409	26126	0	421	6473396
K	830462	4163	85355	0	303	1.91E+07
L	830462	6137	143289	0	465	5.35E+07
M	830462	44	354	1	10	82637
W	830462	1675	16454	0	341	3823710

Age: number of years; Export Status: dummy equal to one for exporters; VA: value added; K: total assets; L: number of workers; M: Material inputs; W: wage bill (including taxes).

Table 9: First stage: *département*

Dep. Var. :	$\Delta Urbanization_{dst}$		$\Delta Location_{idst}$	
	(1)	(2)	(3)	(4)
$Urbanization_{dst-2}$	-0.0083***	-0.0083***	0.0018	0.0018
	(0.0017)	(0.0013)	(0.0027)	(0.0024)
$Location_{idst-2}$	0.0101***	0.0101***	-0.0356***	-0.0356***
	(0.0031)	(0.0029)	(0.0036)	(0.0029)
$\#Firms_{dst-2}$	-0.0133***	-0.0133***	0.0280***	0.0280***
	(0.0048)	(0.0048)	(0.0039)	(0.0033)
...				
Controls	Yes	Yes	Yes	Yes
FEs	i& t	i& t	i& t	i& t
Cluster Level	i & st	i & dst	i & st	i & dst
Observations	460,320	460,320	460,320	460,320

Standard errors in parenthesis: one, two, and three asterisks denote statistical significance at the 10%, 5%, and 1% level, respectively. Regressions include all firm and industry covariates.

Table 10: First stage: employment areas

Dep. Var. :	$\Delta Urbanization_{dst}$		$\Delta Location_{idst}$	
	(1)	(2)	(3)	(4)
$Urbanization_{dst-2}$	-0.0108***	-0.0108***	0.0008	0.0008
	(0.0014)	(0.0012)	(0.0026)	(0.0019)
$Location_{idst-2}$	0.0052**	0.0052**	-0.0026***	-0.0608***
	(00021)	(0.0021)	(0.0012)	(0.0020)
$\#Firms_{dst-2}$	-0.0075*	-0.0075*	-0.0561***	-0.0561***
	(0.0043)	(0.0042)	(0.0033)	(0.0030)
...				
Controls	Yes	Yes	Yes	Yes
FEs	i& t	i& t	i& t	i& t
Cluster Level	i & st	i & dst	i & st	i & dst
Observations	460,320	460,320	460,320	460,320

t-stat in parenthesis: one, two, and three asterisks denote statistical significance at the 10%, 5%, and 1% level, respectively. Regressions include all firm and industry covariates.

Table 11: Evolution of labor Gap by selected period, real euro (thousand)

Dep. Var. :	Labor Gap $\|G^l_{it}\|$		
	(1)	(2)	(3)
1998-2000	-0.048**	-0.041**	-0.041
	(0.021)	(0.021)	(0.076)
2001-2003	0.888***	0.858***	0.858***
	(0.029)	(0.029)	(0.075)
2004-2007	1.723***	1.654***	1.654***
	(0.042)	(0.042)	(0.079)
SIZE: 2^{nd} quintile		-1.334***	-1.334***
		(0.067)	(0.076)
SIZE: 3^{rd} quintile		-2.274***	-2.274***
		(0.093)	(0.116)
SIZE: 4^{th} quintile		-2.927***	-2.927***
		(0.122)	(0.156)
SIZE: 5^{th} quintile		-3.480***	-3.480***
		(0.167)	(0.216)
$Comp_{st}$		0.071***	0.071*
		(0.021)	(0.042)
Exp_{it}		-0.236***	-0.236***
		(0.034)	(0.036)
$ln(age)_{it}$	-0.107	0.291**	0.291*
	(0.127)	(0.127)	(0.155)
$ln(age)^2_{it}$	0.012	-0.050**	-0.050**
	(0.021)	(0.021)	(0.025)
Cluster Level	i	i	i & st
Observations	830,462	830,462	830,462
# of Firms	103,046	103,046	103,046
R-squared (Between)	0.617	0.618	0.618

Standard errors in parenthesis: *, **, *** denote statistical significance at the 10%, 5%, and 1% level, respectively. All regressions include firm fixed effects. Dependent variable: labor gap in real euro. Marginal productivity of labor is computed using Wooldridge's (2009) modification of the Levinsohn-Petrin algorithm which considers capital as a fixed input, labor as a flexible input, raw materials as a proxy for unobserved productivity shocks, and a Cobb-Douglas production function.

Table 12: Agglomeration externality: Cobb-Douglas production function (GMM-IV)

Dep. Var.:	Labor Gap $\lvert G^l_{it} \rvert$			
	(1)	(2)	(3)	(4)
	département		Employment Areas	
$Urbanization_{dst}$	-2.694**	-2.671***	-2.849***	-2.889***
	(1.039)	(0.835)	(0.917)	(0.708)
$Location_{idst}$	-0.112	-0.123	0.032	0.036
	(0.400)	(0.338)	(0.140)	(0.141)
...				
Controls	Yes	Yes	Yes	Yes
FEs	i & t	i & t	i& t	i & t
Cluster Level	i & st	i & dst	i & st	i & dst
Observations	460,320	460,320	460,320	460,320
# of Firms	87,303	87,303	87,303	87,303
First Stage F-test	43.38	59.19	28.41	36.67
Hansen J (p-value)	0.175	0.0377	0.689	0.562

Standard errors in parenthesis: *, **, *** denote statistical significance at the 10%, 5%, and 1% level, respectively. All regressions include firm and year fixed effects along with all the others controls. Dependent variable: labor gap in real euro. Single plant firms only. First stage F test refers to the Kleibergen-Paap F statistics.

Table 13: Robustness to Different Clustering levels (GMM-IV)

Dep. Var.:	Labor Gap $\lvert G^{l}_{it}\rvert$			
	(1)	(2)	(3)	(4)
	Département		Employment Area	
$Urbanization_{dst}$	-2.288***	-2.330***	-2.485***	-2.776***
	(0.863)	(0.882)	(0.641)	(0.636)
$Location_{idst}$	-0.258	-0.087	-0.111	0.033
	(0.343)	(0.367)	(0.137)	(0.143)
...				
Controls	Yes	Yes	Yes	Yes
Production Function	TL	CD	TL	CD
FEs	i & t	i & t	i & t	i & t
Cluster Level	ds	ds	ds	ds
Observations	460,320	460,320	460,320	460,320
First Stage F-test	55.73	55.73	60.42	60.42
Hansen J (p-value)	0.151	0.0439	0.724	0.521

Standard errors in parenthesis: *, **, *** denote statistical significance at the 10%, 5%, and 1% level, respectively. All regressions include firm and year fixed effects along with all the others controls. Dependent variable: labor gap in real euro. Single plant firms only. First stage F test refers to the Kleibergen-Paap F statistics. Note: TL stands for Trans Logarithmic whereas CD for Cobb-Douglas.

Table 14: Robustness to Different Samples (GMM-IV)

Dep. Var.:	Labor Gap $\|G_{it}^{l}\|$			
	(1)	(2)	(3)	(4)
	Départment		Employment Area	
$Urbanization_{dst}$	-1.564	-3.143**	-2.007**	-3.680***
	(1.019)	(1.378)	(0.812)	(0.953)
$Location_{idst}$	-0.263	-0.001	-0.182	-0.004
	(0.412)	(0.390)	(0.129)	(0.154)
...				
Controls	Yes	Yes	Yes	Yes
FEs	i & t	i & t	i & t	i & t
Cluster Level	i	i & dt	i	i & dt
Outliers	$1^{st}pc - 99^{th}pc$	$1^{st}pc - 99^{th}pc$	$1^{st}pc - 99^{th}pc$	$1^{st}pc - 99^{th}pc$
		and $\geq$ 10 workers		and $\geq$ 10 workers
Observations	447,712	213,030	447,712	213,030
# of Firms	73,331	36,781	70,252	36,781
First Stage F-test	43.38	29.74	28.34	31.75
Hansen J (p-value)	0.456	0.893	0.774	0.714

Standard errors in parenthesis: *, **, *** denote statistical significance at the 10%, 5%, and 1% level, respectively. All regressions include firm and year fixed effects along with all the others controls. Dependent variable: labor gap in real euro. Single plant firms only. First stage F test refers to the Kleibergen-Paap F statistics.