Journal of Economic Geography **19** (2019) pp. 251–272
Advance Access Published on 23 March 2018
doi:10.1093/jeg/lby007

Agglomeration economies and firm-level labor misallocation

Lionel Fontagné[*,**,†] *and Gianluca Santoni*[**]

*PSE-University Paris 1 Panthéon-Sorbonne, Centre d'Economie de la Sorbonne, 110 boulevard de l'Hôpital 75013 Paris, France
**CEPII, 20 avenue de Ségur, 75334 Paris cedex 07, France
†Correspondence to: *email* <lionel.fontagne@univ-paris1.fr>

Abstract

A large portion of the productivity differentials among locations is related to density. Firms located in denser areas are more productive (Combes et al., 2012a, *Econometrica*, 80: 2543–2594). We show in this article that resource misallocation is lower in denser areas and argue that misallocation plays an important role in explaining the observed productivity premium of large cities. Using a methodology proposed by Petrin and Sivadasan (2013, *Review of Economics and Statistics*, 95: 286–301), we firstly assess the degree of resource misallocation among firms within sectors for each of the 348 French 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 1994–2007, the average gap at firm level is 9.5 thousand euros. Secondly, we show that firm misallocation is lower in denser employment zones, suggesting that matching mechanisms on the labor market contribute to the productivity premium of agglomerated locations.

Keywords: Misallocation, agglomeration, productivity, firm-level data
JEL classifications: D24, R12, L25, O47
Date submitted: 30 November 2016 **Editorial decision:** 8 February 2018
Date accepted: 1 March 2018

1. 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 also be due 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), technology spillovers (learning) and a denser labor market (matching).[1]

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

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 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 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; for example, US 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% while 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 toward the most efficient firms is conducive to aggregate performance.

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

3 Syverson (2004) reports a TFP ratio of 1.92 for the USA 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 USA 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). US productivity inevitably involves gaps and a degree of misallocation, the distribution is used as the control group.

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 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. (2017) 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 4-fold increase

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 and Development (OECD), EU Commission), universities and think-tanks and non-European central banks (Argentina and Peru) and organizations (US 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 USA shows a correlation of about 50%. The results in Bartelsman et al. (2013) show a higher covariance for European countries, ranging from 15% to 38%, confirming the existence of a sizeable efficiency gap with respect to the US 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 USA and France.

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 article is organized as follows. Section 2 presents the data and descriptive evidence of differences in firm productivity. The methodology described in Section 3 was inspired by Petrin and Sivadasan (2013). Total factor productivity (TFP) estimation strategy is described in Section 3.1. Section 4 computes the value of the labor gaps at both sector (Section 4.1) and firm (Section 4.2) level. In Section 5, we assess the effect of agglomeration economies on the dynamics of labor gaps, controlling for firm characteristics and the endogeneity of density variables. The last section concludes.

2. Data and productivity estimation

Our evaluation of input allocation is performed using firm-level balance sheet data to retrieve 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 our geographical unit of interest: the commuting zone level (*Zone d'emploi*—EZ). We ignore the municipality level, which for our purposes is meaningless. In contrast to alternative zoning, such as the *Département*,[8] commuting zones were defined for statistical purposes and based on the observed distribution of economic activities. The definition of EZ from the National Institute of Statistics (INSEE) read as follow: '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'. Even though the EZs were revised in 2011 based on the 2006 census in the empirical analysis, we use the 1993 definition (based on the 1990 census) which divides the French territory into over 348 nonoverlapping commuting zones.

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, that is, those reporting negative or zero values for our variables of interest, and cleaning the data of potential outliers,[10] we have an

8 An administrative border for local governments introduced in 1789.
9 BRN is the normal tax regime for French firms. In 2007, manufacturing firms declaring under the BRN regime accounted for 94% of overall employment and 96% of manufacturing value added.
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, for example, 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*). 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. The corresponding tables are reported in the Online Appendix.

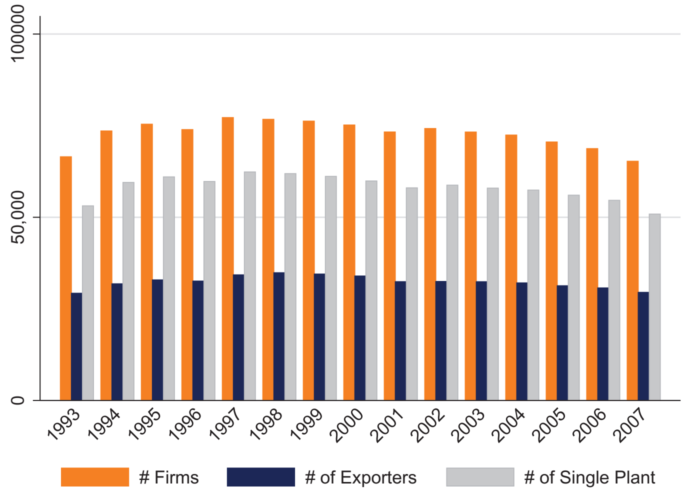

Figure 1. Distribution of firms in the sample, number of exporters and single-plant firms (manufacturing only).

unbalanced panel of 120,198 manufacturing firms, of which 105,267 are single-plant firms (Figure 1).[11] Since we observe financial accounts at the firm (legal unit) level, production functions must be estimated using information on the universe of firms in the sector—including multi-plant firms (i.e. legal units comprising several production units). In contrast, since multi-plant firms may have production units located in different EZs, we must restrict our sample to single-plants when studying the relation between agglomeration and labor misallocation. Provided the productivity premium of exporters repeatedly documented in the empirical international economics literature (see Bernard et al. (2007) and Wagner (2012) for a recent survey), the composition of our sample is an important issue. Figure 1 points however to a stable proportion of exporters over time, suggesting that the empirical evidence discussed below is not driven by sample composition effects.

Before turning to more sophisticated analysis we investigate whether firms in denser (more agglomerated) areas are more productive overall, even when controlling for demand-driven scale economies. Indeed, this comparison has to be performed for firms located in one EZ only. 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 EZs below and above the median agglomeration (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 mitigate simultaneity bias, the measure of agglomeration is based on the first available year in

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.

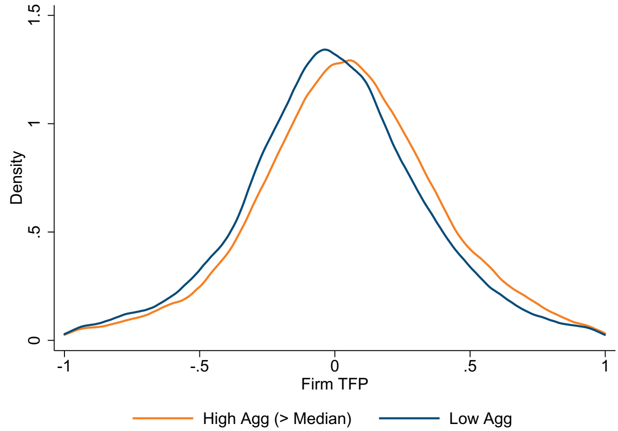

Figure 2. Urbanization and productivity in EZs below and above the median agglomeration, single-plant firms (manufacturing only year 2000).

the sample, that is, 1994. This premium is 6.4%, significant at the 1% level. Although we do not include any controls, the difference is clear.

3. Evaluating misallocation 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 the currently available resources would allow due to their inefficient distribution only. 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 is 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 Sivadasan (2013)'s methodology, used originally to evaluate the impact of a change in labor market regulation in Chile. 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. In this article, we observe balance sheets of legal units, which can be single- plant or multi-plant firms. The concept of firm-specific 'misallocation' refers to the gap between the value of the marginal product and the marginal input price. Taking the sector-level production function (estimated for all firms in the sector) as a benchmark, 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 this 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 next step of our analysis is to address the impact of location on gaps, and to do so we restrict the analysis to single-plant legal units in order to have a one-to-one relationship between firm and location.

The underlying economic intuition is that in a context of perfect competition in the factor market, 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. Interestingly, considering the value of the marginal product of labor is consistent with perfect competition, oligopolistic or monopolistic competition in the product market (see e.g. Hsieh and Klenow (2009) for the case of monopolistic competition with heterogenous firms).

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:

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

Where q_{it} denotes value added, l_{it} is the number of employees and k_{it} is the fixed capital stock. All series are represented in logs. As Loecker et al. (2012) note, the translog reduces to a Cobb–Douglas when we drop the squared and interaction terms, for example, exclude the control for relative inputs intensity. Our results are robust to using a Cobb–Douglas production function.[12] 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: $\phi_{\mathrm{it}}^l = \beta_l + 2\beta_{\mathrm{ll}} l_{\mathrm{it}} + \beta_{\mathrm{lk}} k_{\mathrm{it}}$.

The error term has two components: ω_{it} represents a Hicks-neutral productivity shock (observed by the firm but not by the econometrician), and $\varepsilon_{\mathrm{it}}$ 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 and other demand factors (see Section 3.1 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.[13] 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 because our research question is how density of employment zones impact the matching on the labor market.[14] The marginal product of labor is given as the marginal increment in revenue per unit change in labor:

$$VMP_{\mathrm{it}}^L = \frac{\partial P_{\mathrm{it}} * Q_{\mathrm{it}}}{\partial L} = \phi_{\mathrm{it}}^l \frac{P_{\mathrm{it}} * Q_{\mathrm{it}}}{L_{\mathrm{it}}} \tag{2}$$

12 See the results in the Online Appendix.
13 Or reaching the same level of output using less input.
14 It can be generalized to any input.

Since output prices at firm level are generally not available (or available only for a subsample of the surveyed firms) and using industry prices as proxy may worsen measurement errors (Foster et al., 2008), we only control for price changes over time at the industry level by including year dummies in Equation (1). As noted by Petrin and Sivadasan (2013), marginal products of inputs, that is, ϕ_{it}^{z}, 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) when controlling for firm fixed effects as in Equation (7).

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

$$G_{\mathrm{it}}^{L} = VMP_{\mathrm{it}}^{L} - w_{\mathrm{it}} \tag{3}$$

Where w_{it} represents the wage of the marginal worker in firm i.[15] To ease comparability over time, the value of G_{it}^{L} is deflated using the consumer price index (CPI).[16] The absolute value of G_{it}^{L} 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 and taxes 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.

3.1. Productivity estimation

The first step toward assessing the input gaps is to compute firm-level TFP. Our measure of TFP is computed based on Wooldridge (2009)'s implementation of Levinsohn and Petrin (2003)'s algorithm, using material inputs to proxy for technology shocks, and considering labor as a semi-flexible input and capital as fixed.[17]

However, since we are interested in capturing the impact of employment zone density on labor gaps, we need to control for any additional impact of agglomeration on productivity. As Saito and Gopinath (2009) show, location attributes, and in particular, demand-driven economies of scale, may affect productivity estimates. In estimating plant-level productivity for Chilean food industry firms, Saito and Gopinath (2009) introduced three measure of market characteristics in the production function, that is,

15 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.

16 Results are robust to the use of the GDP deflator instead of CPI.

17 The semi-parametric estimator proposed by Levinsohn and Petrin (2003) extends the methodology in Olley and Pakes (1996), which suggests using investment 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. This is the main reason why estimators that ignore this correlation produce inconsistent results, for example, OLS. 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).

regional industry share, employment concentration and regional population; they find that only regional population (a proxy for demand-driven economies of scale) has a positive and significant effect on plant productivity.[18] As De Loecker (2011) suggests, failing to control for prices and demand shifters potentially might bias the production function coefficients. In the empirical application, we follow De Loecker (2011) and to proxy for local demand, include the municipality population in year 1968 as an additional instrument in the productivity estimation. In all the empirical results reported in the article, marginal labor productivity—used to build the gaps—is conditional on demand-driven economies of scale that might affect the firm's input demand choice, and thus, the production function estimates.[19]

Wooldridge suggests implementing Levinsohn and Petrin's methodology in a general method of moments (GMM) framework that takes account of potential contemporaneous error correlation among the two stages, as well as heteroskedasticity and serial correlation. It is also robust to the critique by Ackerberg et al. (2015).[20] Note that our findings are robust to different estimation methodologies, namely semi-parametric two-stage estimation as well as a (firm) fixed-effect estimator.[21]

Figure 3 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 7.1% 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 non-negligible redistribution of firms toward 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 0.86, meaning that for a given amount of inputs, the most efficient firms were able to reach a level of production 137% higher than the least efficient firms.[22] In 2007, given the average increase in productivity, the interquartile ratios increased to 0.96 (i.e. 160%), 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.

4. 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. Gaps are deflated by CPI for the sake of comparison over time.

18 We thank an anonymous referee for the suggestion.
19 All the empirical results reported in the article are robust to the exclusion of this additional instrument in the production function estimations; the corresponding tables are available upon request.
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 $\exp(0.86) - 1 = 137$.

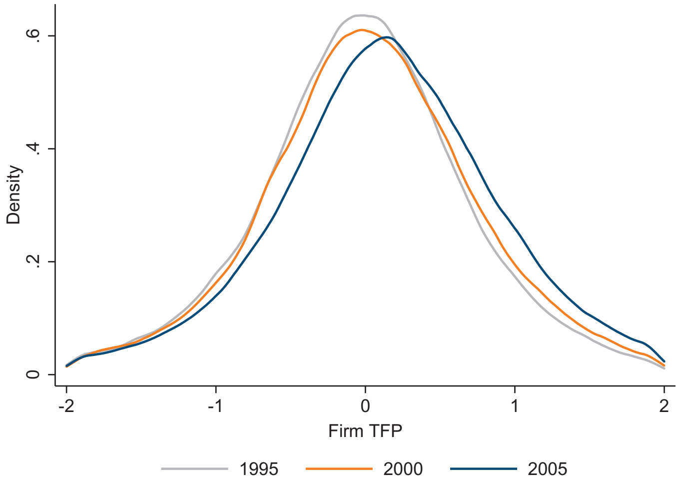

Figure 3. Manufacturing firms' productivity distribution, ω_{it}.
Note: The graph reports the distribution of manufacturing firm TFP for selected years. Pooled distribution is standardized (over the whole period) to have zero mean and standard deviation equal to 1. The three distributions are statistically different at the 1% confidence level.

4.1. 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 in terms of input cost share. Estimated returns to scale are generally slightly below unity (on average 0.94); 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 (2) and (3). The main results for the labor return to cost wedge are reported in Table 1.

For a given sector s at time t, the mean absolute labor gap is defined as follows $Gap_s^{Abs} = \frac{\Sigma_{i \in s} |G_i^L|}{N_s}$. It measures the distance from the social optimum allocation[23] where each firm is equating marginal revenue equal to marginal costs and there is no friction in the input markets. We will investigate below whether a better matching on a thicker local market helps reducing such gaps, conditional on regulatory distortions common to all locations in a single market. For the whole manufacturing sector over the period 1994–2007, this figure is 9502 euros per firm,[24] with dispersion relatively high both between and also within industries, as shown by a coefficient of variation (CV) above 1 for most of the sectors. The mean absolute labor gap varies much across sectors, ranging from around 6152 euros (resp. 6763 euros) in the wood (resp. food) industries, to 17,863 euros in the pharmaceutical industries. Such differences suggest that the quality of the match on the local labor market can be particularly important in certain activities.

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.

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	Pos %	Obs
Basic metals	0.652	0.223	0.875	0.064	8.994	1.023	24.0	10,413
Beverages	0.646	0.377	1.023	0.065	17.234	0.996	65.5	11,529
Chemicals	0.648	0.232	0.880	0.025	12.456	1.066	38.4	25,642
Computer and electronics	0.678	0.225	0.904	0.041	10.991	0.986	24.0	33,042
Electrical equipment	0.655	0.290	0.945	0.070	9.256	0.969	20.3	21,727
Fabricated metal	0.673	0.264	0.937	0.025	8.069	0.919	20.2	175,789
Food products	0.639	0.252	0.891	0.053	6.763	1.048	27.9	160,419
Furniture	0.655	0.266	0.921	0.044	8.136	0.837	14.6	35,982
Leather products	0.752	0.308	1.060	0.041	6.982	1.269	30.0	10,750
Machinery and equipment	0.704	0.235	0.940	0.036	9.192	1.046	27.9	66,868
Motor vehicles	0.686	0.204	0.890	0.049	8.220	1.104	24.1	17,245
Nonmetallic products	0.628	0.318	0.947	0.047	9.148	1.012	24.5	38,867
Other manufacturing	0.689	0.303	0.992	0.029	9.272	1.039	29.5	45,268
Other transport	0.694	0.217	0.910	0.055	8.830	1.071	28.0	7750
Paper products	0.656	0.259	0.915	0.027	8.784	1.008	26.8	18,726
Pharmaceutical	0.494	0.275	0.769	0.058	17.863	0.846	23.4	5627
Printing and recording	0.690	0.235	0.925	0.024	8.963	0.964	20.1	81,066
Repair and installation of machinery	0.729	0.197	0.926	0.035	7.682	1.088	26.7	91,306
Rubber and plastic	0.649	0.282	0.931	0.058	8.226	1.026	27.8	46,756
Textiles	0.673	0.332	1.005	0.076	8.521	1.104	26.1	31,342
Wearing apparel	0.730	0.316	1.046	0.074	9.314	1.156	31.0	41,709
Wood products	0.691	0.289	0.980	0.029	6.152	1.134	30.1	43,173
Overall	0.669	0.268	0.937	0.047	9.502	1.032	27.8	1,020,996

If we look at the sample data last year, we observe that a quarter (25.2%) of the manufacturing firms in the sample report a positive wedge between the marginal return and cost of labor (see Table 2).[25] The sign of the gap is meaningful since it helps differentiating the variability from the direction of firm-level misallocation. Recall that resource misallocation implies that more efficient firms tend to be smaller than their optimal size: the marginal revenue of labor is above its cost and the gap is positive because some frictions in the input market prevent these firms from expanding their activity. The average positive gap is roughly 34% higher than the negative counterpart, and the overall distribution for positive wedges seems to be relatively more right skewed with respect to the negative wedges. Assuming an average labor cost of 40,000 euros per year in 2007 in France,[26] an average negative wedge of 8671 euros implies that the marginal return from labor is smaller than the cost of around 2.5 months of salary. Notice that there is a relatively high dispersion in the data since the CV is 1.03. On the other hand, the value produced by the marginal worker is higher by almost 11,599 euros than its cost when a positive wedge is observed (with a CV of 1.21).

25 The share of firms with positive gaps over the whole period is 27.8%.

26 The average labor cost in 2007 at 2005 prices for full time workers—excluding civil servants—was 40,204 euros, including both salary and tax allowances. Source: INSEE-DADS.

Table 2. Labor gap decomposition, year 2007

	$\lvert G^L_{\mathrm{it}}\rvert$	$G^L_{\mathrm{it}} > 0$	$G^L_{\mathrm{it}} < 0$
Number of firms	65,288	16,420	48,868
Share (%)	100	25.2	74.8
Mean	9.407	11.599	8.671
SD	9.588	14.031	7.384
10%	1.517	0.830	1.959
Median	6.834	6.164	6.942
90%	19.336	31.238	16.826

4.2. 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 the current subsection 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, especially in the case of large firms that rely on a large pool of internal competencies. Then, we will investigate in the next section whether firms in denser areas exhibit lower labor gaps, controlling for firm characteristics and accounting for the nonrandom location choices of workers, firms and wage or productivity shocks. 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:

$$Y_{\mathrm{it}} = \alpha_0 + \delta_1 + \delta_2 + \delta_3 + \Gamma_{\mathrm{it}}\beta + \xi_i + \varepsilon_{\mathrm{it}} \tag{4}$$

where Y_{it} is the absolute value of the labor gap, $|G^L_{it}|$. The time evolution of the dependent variable is accounted for by three subperiod dummies: δ_1 for the years 1998–2000, δ_2 for 2001–2003 and δ_3 for 2004–2007. The constant α_0 captures the reference period gap value. The vector Γ_{it} includes a set of firm 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 ($Comp_{\mathrm{st}}$).[28] Finally, ξ_i are firm fixed effects to control for unobserved heterogeneity and $\varepsilon_{\mathrm{it}}$ is an idiosyncratic shock. 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,[29] which requires controlling in the empirical analysis for the firm's pricing strategy. Indeed, markups are positively correlated with firm productivity and export participation and

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 $ln(1/HH)_{\mathrm{st}}$, where HH is an Herfindahl–Hirschman index of sales concentration by sector s and year t.

29 A potential source of concern could be that we do not observe firm-level prices. In the presence of markups the production function estimates may be biased downward; such bias emerges if the difference between the firm and the industry price is correlated systematically with the input choice (De Loecker, 2011; Gopinath et al., 2017).

Table 3. Evolution of labor gap by selected period, real euro (thousand)

Dep. Var.:	Absolute labor gap $\lvert G^L_{it} \rvert$			
	(1)	(2)	(3)	(4)
1998–2000	0.487***	0.371***	0.405***	0.407***
	(0.073)	(0.074)	(0.078)	(0.078)
2001–2003	1.408***	1.176***	1.240***	1.238***
	(0.114)	(0.117)	(0.116)	(0.116)
2004–2007	2.224***	1.867***	1.934***	1.925***
	(0.082)	(0.131)	(0.137)	(0.138)
$ln(age)_{it}$		−1.192***	−0.869***	−0.865***
		(0.081)	(0.085)	(0.085)
$ln(age)^2_{it}$		0.389***	0.368***	0.368***
		(0.044)	(0.045)	(0.045)
Size: 2nd quintile			−1.517***	−1.507***
			(0.074)	(0.074)
Size: 3rd quintile			−2.640***	−2.618***
			(0.104)	(0.103)
Size: 4th quintile			−3.489***	−3.456***
			(0.139)	(0.139)
Size: 5th quintile			−4.291***	−4.243***
			(0.188)	(0.188)
Exp_{it}				−0.227***
				(0.031)
$Comp_{st}$				0.031
				(0.037)
Fixed effects	i	i	i	i
Cluster level	i and st	i and st	i and st	i and st
Observations	1,020, 996	1,020,996	1,020,996	1,020,996
Number of firms	120,198	120,198	120,198	120,198
R-squared	0.559	0.560	0.562	0.562

Notes: Standard errors, clustered at the firm i and sector–year st level, in parenthesis. *, **, *** Statistical significance at the 10, 5 and 1 levels, respectively. Dependent variable: absolute labor gap in real euro (base year 2005).

negatively related to the toughness of competition (Bellone et al., 2016). The set of controls included in the vector Γ_{it} controls directly for unobserved firm prices and variation in demand.

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 subperiod 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 2224 euros higher compared to the reference period (1994–1997). In columns (2)–(4), we add the covariates specified above to proxy for firms' pricing strategy and the estimation results confirm the time evolution of the gap. In all the columns of Table 3, standard errors are double clustered at firm level (in order to deal with serial correlation) and at the sector–year level (to control for cross-sectional dependence). Baseline results confirm that the average gap is increasing

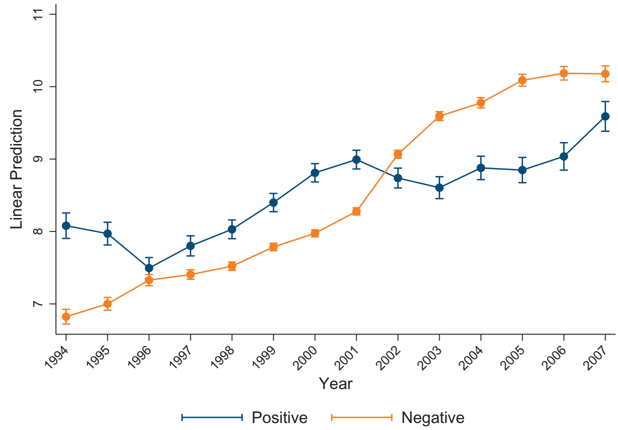

Figure 4. Average labor gap conditional on firm characteristics.

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.[30]

The evolution over time of the labor gap for an average manufacturing firm with positive or negative values is depicted in Figure 4.[31] The dynamics of the negative and positive gaps are different. A sharp increase in the average negative gap is observed from 2001 onward. 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. Different (unobserved) regulations constraining firms below their optimal size may have played an increasing role.

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'). 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.[32] The main effect of this increasing regulation is an increase in labor costs which may induce resource misallocation (see Garicano et al., 2016) and potentially could affect our results. Column (4) shows that

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

31 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 (4).

32 Above this threshold firms are expected also to establish a 'social plan' for more than nine employees are laid off at the same time, to show that the firm owner has tried to find alternative employment for those being dismissed.

Table 4. Evolution of labor gap sample sensitivity

Dep. Var.:	Absolute labor gap $\|G_{it}^{L}\|$				
	Restricted sample	Small firms	≤49 workers	Single plant	Single plant and $G_{it}^{L} < 0$
1998–2000	0.626***	0.277***	0.339***	0.325***	0.195**
	(0.087)	(0.081)	(0.078)	(0.077)	(0.080)
2001–2003	1.431***	1.077***	1.172***	1.104***	1.172***
	(0.119)	(0.130)	(0.124)	(0.118)	(0.160)
2004–2007	2.239***	1.641***	1.809***	1.691***	1.931***
	(0.135)	(0.162)	(0.148)	(0.146)	(0.167)
Controls	Yes	Yes	Yes	Yes	Yes
Fixed effects	i	i	i	i	i
Cluster level	i and st	i and st	i and st	i and st	i and st
Observations	305,583	708,860	891,271	812,307	592,515
Number of firms	36,266	94,060	109,594	102,344	83,097
R-squared	0.621	0.567	0.563	0.565	0.527

Notes: Standard errors, clustered at the firm i and sector–year st level, in parenthesis. *, **, *** Statistical significance at the 10, 5 and 1% levels, respectively. Dependent variable: absolute labor gap in real euro (base year 2005).

firms which remain below the 50-worker threshold, other things being equal, show the same time pattern as the whole sample.

The rationale for keeping the whole sample of firms for the estimation of the production function has been stressed above. The discussion of labor gaps consistently considered the whole sample. Before turning to agglomeration economies' evidences by single-location firms, it is worth focusing on the dynamics of labor gaps for single-plant firms. This is done in the last two columns of Table 4. We observe no qualitative difference in the evolutions. In quantitative terms, the adverse evolution is reinforced if we stick to single-plant firms exhibiting negative gaps (last column).

5. Agglomeration economies and labor misallocation

We now address our central argument and enrich the baseline specification in Equation (4) to test for the effect of agglomeration economies on return-cost wedges. Comparing the empirical firm productivity distribution across high- versus low-agglomerated 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 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 more specialized areas the thicker labor market also affects firm resource allocation efficiency, that is, return to cost wedge. To control for the size of the local labor market and agglomeration externalities, we include in the estimated equation a measure of the economic specialization of EZ and enrich the set of controls. Our main variable of interest, $Location_{zst}$, measures the number of employees in an EZ z by sector s and year t.[33] The estimated equation becomes:

$$ln(Y_{it}) = \beta_1 ln(Location_{zst}) + \Gamma_{it}\beta + X_{zst}\beta + \tau_{zs} + \tau_{st} + \varepsilon_{it} \quad (5)$$

where $ln(Y_{it})$ is the log of the absolute labor gap, $|G^L_{it}|$. As additional controls in Equation (5) we include the average productivity by EZ, sector and year (zst) in order to control for the effects of agglomeration economies on productivity and the share of exporters around firm i to further control for agglomeration externalities; we add to the vector of firm level covariates a dummy variable for firms with positive gaps. Fixed effects by EZ, τ_{zs}, controls for any unobservable, time-invariant, characteristic of the local labor market whereas sector–year fixed effects, τ_{st}, capture any change in the degree of competition and regulation at the industry level. In order to control for EZ internal geography, which may affect the distribution of firms within the EZ, we also include a set of controls at the municipality level: the distance of the municipality to the center of the EZ,[34] to proxy for differential land prices within EZ (see Combes et al., 2016) and a dummy for coastal and for mountainous municipalities.[35] Note that we do not introduce any firm fixed effect: as pointed out in Combes et al. (2011), estimating Equation (5) with firm fixed effect would imply that β_1 is identified only for a selected subsample of firms, that is, those that change location over time. In order to control for firms characteristics, we follow a two-step alternative estimation procedure detailed below. Finally, to limit measurement errors stemming from multi-plant (potentially multi-location) enterprises, we recall to restrict the sample to single-plant firms. Table 5 reports the descriptive statistics for the variables included in the regression.

Baseline ordinary least squares (OLS) results are reported in Table 6 and suggest that on average, denser areas are associated with lower gaps, a result robust to the inclusion of the whole set of controls (column (4)). In columns (5) and (6), we opt for a more conservative correction of the standard errors using a two-way cluster at the EZ and sector–year; furthermore, in column (6) we also test the robustness of our results to a different set of fixed effects. Besides the correction of TFP,[36] the inclusion of location

33 $Location_{zst} = ln(X_{zst})$, where X represents the number of number of workers; our findings are robust to the use of the number of firms as a measure of location, the additional tables are available upon request.

34 The EZ centroid is identified as the weighted average of corresponding municipalities; the weights are based on 1968 population.

35 Municipalities are the lowest tier of local administrations and represent a finer geographical unit of analysis nested within EZs.

36 Note that estimation of TFP is purged from economies of scale; consequently, the gaps that we explain here also take into account local market size.

Table 5. Descriptive statistics

Variable	Mean	SD	10%	Median	90%
$\ln(\lvert G^L_{\mathrm{it}}\rvert)$	1.65	1.09	0.25	1.79	2.85
$\ln(Location_{\mathrm{zst}})$—Workers	6.10	1.49	4.14	6.22	8.00
Natural disasters	0.35	2.11	0.00	0.04	0.44
Variance decomposition	$ln(Y_{it})$ R^2				
Fixed effecs:					
EZ-sector *zs*	0.040				
EZ-sector and sector–Year zs and sy	0.048				
EZ–sector–year zst	0.078				
Firm zsy	0.413				

Note: Estimation sample: Obs = 827,379. Number of firms = 117,433.

Table 6. Agglomeration and labor gap, OLS (firm level)

Dep. Var.:	$\ln(\lvert G^L_{\mathrm{it}}\rvert)$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(Location_{\mathrm{zst}})$	−0.033***	−0.038***	−0.040***	−0.040***	−0.040***	−0.019***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.006)
Controls:						
Firm, it		Yes	Yes	Yes	Yes	Yes
EZ, zst			Yes	Yes	Yes	Yes
Geography *a*				Yes	Yes	Yes
Fixed effects	zs and st	zs and st	zs and st	zs and st	zs and st	z and st
Cluster level	z	z	z	z	z and st	z and st
Observations	827,379	827,379	827,379	827,379	827,379	827,379
Number of firms	117,433	117,433	117,433	117,433	117,433	117,433
R-squared	0.049	0.054	0.054	0.054	0.054	0.037

Notes: Standard errors in parenthesis; *, **, *** Statistical significance at the 10, 5 and 1% levels, respectively. The dependent variable is the log of the absolute labor gap. $Location_{\mathrm{zst}}$ measures the number of workers (L) in an EZ (z), sector (s), year (t). Firms controls include: age, age squared, production quintile dummies, dummy for firms with negative Gaps. EZ controls include: the average productivity (zst), share of exporters (zst). Geography includes: the distance of municipality (a) from the EZ centroid, a dummy for coastal municipalities and a dummy for mountainous municipalities.

characteristics (i.e. EZ fixed effects and internal geography) in Equation (5) should mitigate the risk of omitted variable bias; however, nonrandom location choices of workers, firms as well as nonrandom distribution of shocks (both wage and productivity shocks) could still affect the identification of β_1.

In order to deal with endogeneity, we rely on a Bartik (1991) type instrument for $ln(Location_{\mathrm{zst}})$, using the past EZ's sectoral share and the trend in employment in the rest of the country to build the expected number of workers (or firms) in a location–sector–year cell, zst. We start by running the following poisson regression:

$$Location_{\text{zst}} = \gamma_1 Share_{\text{zs}}^{1968} + \gamma_2 ln(emp_{\text{st}-1}) + \gamma_3 Share_{\text{zs}}^{1968} * ln(emp_{\text{st}-1}) + \gamma_z^T + \varepsilon_{\text{zst}} \quad (6)$$

where $Share_{\text{zs}}^{1968}$ represents the share of workers in a given location–sector at the 1968 population census,[37] $ln(emp_{\text{st}-1})$ is the total employment in sector s at time $t-1$ in the rest of the country, excluding EZ z and the other EZs in the same Region[38] of location z; $Share_{\text{zs}}^{1968} * ln(emp_{\text{st}-1})$ is an interaction term and γ_z^T is a set of location–period fixed effects.[39] The predicted values from Equation (6), $\widehat{Location}_{zst}$, gives the distribution of the economic activity implied by the initial distribution (1968) and an aggregate sectoral trend, but orthogonal to local economic conditions (we exclude $\hat{\gamma}_z^T$ from the prediction). As an additional instrument, we also include the number of records of natural disasters by municipality and year on the government official journal as a measure of exogenous local shocks.[40]

Instrumental variables results are reported in Table 7. The estimated coefficients of *Location* are larger in magnitude and statistically significant, confirming our previous findings for the effect of specialization on the allocation of labor at firm level. In terms of the magnitude, doubling the degree of specialization in an EZ will decrease the gap by 7.4% on average. A simple back of the envelope calculation shows that moving a firm from the 10th to the 90th percentile of *Location* will reduce labor gap of about 0.285, that is, 10.9% of the corresponding 90th–10th percentile *Gap* differential.[41] The reported first stage F-test strongly rejects weak instruments, whereas the Hansen p-value supports the validity of our instruments, pointing to a causal interpretation of the effect of specialization on labor allocation. However, as shown in Table 5, most of the variation in $ln(Y_{it})$ is accounted by the firm fixed effect (roughly 41%). In order to control for firm characteristics and productivity sorting into larger cities, we follow a two-step approach as suggested by Combes et al. (2011). In the first step, we estimate the following equation:

$$ln(Y_{\text{it}}) = \Gamma_{\text{it}}\beta + \phi_{\text{zsT}} + \phi_{\text{iz}(i)} + \varepsilon_{\text{it}} \quad (7)$$

where the vector Γ_{it} includes the firm-level controls and $\phi_{\text{iz}(i)}$ for firm-by-EZ fixed effects. The fixed effect ϕ_{zsT} will capture the average gap by location–sector–period,

37 We use the first census wave available, we also tested 1975 and 1982 waves and results are not significantly affected. Data Source: Population Census 1968: sampling 1/4, INSEE, available at ADISP-CMH. We are deeply indebted to a referee for suggesting this instrumentation strategy. See also Verdugo (2015) for an alternative instrument of location choices (of immigrants), namely the destructions during the Second World War.

38 The Regions are the first tier of local government, during the estimation period the country was divided into 13 nonoverlapping Regions—EZ are nested within Region boundaries.

39 Due to the high number of fixed effects with 348 EZ for as long as 14 years, we divide our sample into three subperiods ($T=3$); we also test with different period lengths ($T=5$ and $T=7$); results are not affected by the threshold.

40 In case of an extreme natural event, such as flood, wildfire, landslide or other meteorological phenomena, the majors of the affected municipalities may ask to the central government for the declaration of 'natural disaster', the publication of the official declaration is mandatory for the financial compensations. We use as instrument the number of declarations in all municipalities of a given EZ in year $t-10$. Source: *Ministère de la Transition écologique et solidaire*, http://www.georisques.gouv.fr/acces-aux-donnees-gaspar.

41 Consider that the (log) difference between the 90th and the 10th percentile of *Location* is 3.85 whereas the corresponding Gap (log) difference is 2.6, see Table 5.

Table 7. Agglomeration and labor gap, IV (firm level)

Dep. Var.:	$\ln(\lvert G^{L}_{it}\rvert)$				
	(1)	(2)	(3)	(4)	(5)
$\ln(Location_{zst})$	−0.095***	−0.092***	−0.096***	−0.090***	−0.074***
	(0.018)	(0.018)	(0.018)	(0.023)	(0.022)
Controls:					
Firm, it		Yes	Yes	Yes	Yes
EZ, zst		Yes	Yes	Yes	Yes
Geography *a*		Yes	Yes	Yes	Yes
IV: Equation (6)	1.307***	1.336***	1.338***	1.339***	1.211***
IV: Natural disasters			−0.012***	−0.012***	−0.010***
F-test	138.1	122.9	254.4	57.62	46.35
Hansen J			0.165	0.325	0.975
Fixed effects	zs and st	zs and st	zs and st	zs and st	z and st
Cluster level	z	z	z	z and st	z and st
Observations	827,379	827,379	827,379	827,379	827,379
R-squared	0.047	0.053	0.053	0.053	0.035

Notes: Standard errors in parenthesis; *, **, *** Statistical significance at the 10, 5 and 1% levels, respectively. The dependent variable is the log of the absolute labor gap. $Location_{zst}$ measures the number of workers in an EZ (z), sector (*s*), year (*t*). Firms controls includes: age, age squared, production quintile dummies, dummy for firms with negative Gaps. EZ controls includes: the average productivity (zst), share of exporters (zst). Geography includes: the distance of municipality (*a*) from the EZ centroid, a dummy for coastal municipality, a dummy for mountainous municipality.

conditional on firm characteristics. The estimated effects $\hat{\phi}_{zsT}$ are then regressed on our measure of agglomeration.[42] Due to the high number of fixed effects to be estimated in the first step—equation (7)—we divide our sample in tree subperiods *T*, in the second step $Location_{zst0}$ refers to value of the variable at the beginning of each period. Results are reported in Table 7 and largely confirm the previous empirical evidence both with and without instrumenting for $Location_{zst0}$.[43]

6. 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 good 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 (single-plant) firms within Employment Zones is observed which provides information on the degree of

42 Since the dependent variable is estimated we perform our regression using both a Weighted Least Square (using the inverse of the standard errors of $\hat{\phi}_{zsT}$ as weights) and a Feasible Generalized Least Squares where we correct the estimation following the same procedure as Combes et al. (2008).

43 Note that the set of instruments in the bottom panel of Table 7 is the same as in the firm-level regression.

Table 8. Agglomeration and labor gap (aggregate level)

Dep. Var.:	$\hat{\phi}_{zsT}$				
	(1)	(2)	(3)	(4)	(5)
OLS					
ln($Location_{zst0}$)	−0.028***	−0.064***	−0.035***	−0.035***	−0.021**
	(0.006)	(0.008)	(0.006)	(0.011)	(0.010)
IV					
ln($Location_{zst0}$)	−0.090***	−0.138***	−0.101***	−0.097**	−0.112***
	(0.013)	(0.017)	(0.013)	(0.035)	(0.037)
F-test	447.5	328.5	433.0	122.9	107.4
Hansen J	0.181	0.442	0.202	0.177	0.291
Correction	None	WLS	FGLS	FGLS	FGLS
Fixed effects	zs and st	zs and st	zs and st	zs and st	*z* and st
Cluster level	*z*	*z*	*z*	*z* and st	*z* and st
Observations	17,744	17,744	17,744	17,744	17,744
R-squared (OLS)	0.334	0.472	0.335	0.335	0.079

Notes: Standard errors in parenthesis; *, **, *** Statistical significance at the 10, 5 and 1% levels, respectively. The dependent variable is the estimated EZ–sector–period fixed effects from Equation (7). The set of instruments for the IV regression includes $\widehat{Location}_{zst}$—predicted from Equation (6)—plus the number of Natural Disasters. Since the dependent variable is estimated column (2) weights observations by the inverse of the standard error of $\hat{\phi}_{zsT}$; columns (3)–(5) report the results of a FGLS following Combes et al. (2008). $Location_{zst}$ measures the number of workers in an EZ (z), sector (s), year (t). Firms controls in the ancillary regression of $\hat{\phi}_{zsT}$ include: age, age-squared, production quintile dummies and a dummy for firms with negative gaps.

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 9500 euros.

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. Specialization *at the Employment Zone*, as measured by the number of firms (or employees) by zone, sector and year, is associated with a lower average labor gap. A doubling of the degree of specialization decreases the gap by 5% on average, suggesting that such matching is playing a role in determining the productivity advantage of denser areas.

Supplementary material

Supplementary data for this article are available at *Journal of Economic Geography* online.

Acknowledgements

We acknowledge support from the Bank of France at earlier stages of this project. We are grateful to Claire Lelarge, Thierry Mayer and two anonymous referees for insightful suggestions. We are also indebted to Irene Iodice for excellent research assistance. The views expressed in this publication are the authors' alone, and do not necessarily reflect the views of the European Commission.

Funding

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).

References

Ackerberg, D. A., Caves, K., Frazer, G. (2015) Identification properties of recent production function estimators. *Econometrica*, 83: 2411–2451.

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

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

Bartik, T. J. (1991) *Who Benefits from State and Local Economic Development Policies*? Number wbsle in Books from Upjohn Press. W.E. Upjohn Institute for Employment Research.

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

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

Bento, P., 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., Schott, P. K. (2007) Firms in international trade. *The Journal of Economic Perspectives*, 21: 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.

Combes, P.-P., Duranton, G., Gobillon, L. (2008) Spatial wage disparities: sorting matters! *Journal of Urban Economics*, 63: 723–742.

Combes, P.-P., Duranton, G., Gobillon, L. (2011) The identification of agglomeration economies. *Journal of Economic Geography*, 11: 253–266.

Combes, P.-P., Duranton, G., Gobillon, L. (2016) *The Costs of Agglomeration: House and Land Prices in French Cities*. Mimeo.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., Roux, S. (2012a) The productivity advantages of large cities: distinguishing agglomeration from firm selection. *Econometrica*, 80: 2543–2594.

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

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

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

De Loecker, J. (2011) Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica*, 79: 1407–1451.

Dhingra, S., Morrow, J. (2014) Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity. Mimeo.

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

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

Foster, L., Haltiwanger, J., Syverson, C. (2008) Reallocation, firm turnover, and efficiency: selection on productivity or profitability? *American Economic Review*, 98: 394–425.

Garicano, L., LeLarge, C., Reenen, J. V. (2016) Firm Size Distortions and the Productivity Distribution: Evidence from France. Working Paper 18841, National Bureau of Economic Research.

Gopinath, G., Kalemli-Özcan, S., Karabarbounis, L., Villegas-Sanchez, C. (2017) Capital allocation and productivity in South Europe. *The Quarterly Journal of Economics*, 132: 1915–1967.

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

Hsieh, C.-T., Klenow, P. J. (2009) Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124: 1403–1448.

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

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

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

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

Petrin, A., Sivadasan, J. (2013) Estimating lost output from allocative inefficiency, with an application to Chile and firing costs. *Review of Economics and Statistics*, 95: 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., Santaeulalia-Llopis, R. (2015) *Land Misallocation and Productivity*. Working Papers tecipa-533, University of Toronto, Department of Economics.

Saito, H., Gopinath, M. (2009) Plants' self-selection, agglomeration economies and regional productivity in Chile. *Journal of Economic Geography*, 9: 539–558.

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

Syverson, C. (2011) What determines productivity? *Journal of Economic Literature*, 49: 326–365.

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: 529–569.

Verdugo, G. (2015) Public housing magnets: public housing supply and immigrants location choices. *Journal of Economic Geography*, 16: 237–265.

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

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