

Combes, Pierre-Philippe; Duranton, Gilles; Gobillon, Laurent; Puga, Diego;
Roux, Sébastien

Working Paper

The productivity advantages of large cities: Distinguishing agglomeration from firm selection

IZA Discussion Papers, No. 6502

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Combes, Pierre-Philippe; Duranton, Gilles; Gobillon, Laurent; Puga, Diego; Roux, Sébastien (2012) : The productivity advantages of large cities: Distinguishing agglomeration from firm selection, IZA Discussion Papers, No. 6502, Institute for the Study of Labor (IZA), Bonn,
<https://nbn-resolving.de/urn:nbn:de:101:1-2012080812474>

This Version is available at:

<https://hdl.handle.net/10419/58784>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

IZA DP No. 6502

The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection

Pierre-Philippe Combes
Gilles Duranton
Laurent Gobillon
Diego Puga
Sébastien Roux

April 2012

The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection

Pierre-Philippe Combes

Aix-Marseille School of Economics and CEPR

Gilles Duranton

University of Toronto and CEPR

Laurent Gobillon

INED, PSE, CEPR and IZA

Diego Puga

IMDEA Social Sciences Institute and CEPR

Sébastien Roux

CREST (INSEE)

Discussion Paper No. 6502

April 2012

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection^{*}

Firms are more productive on average in larger cities. Two main explanations have been offered: firm selection (larger cities toughen competition, allowing only the most productive to survive) and agglomeration economies (larger cities promote interactions that increase productivity), possibly reinforced by localised natural advantage. To distinguish between them, we nest a generalised version of a tractable firm selection model and a standard model of agglomeration. Stronger selection in larger cities left-truncates the productivity distribution whereas stronger agglomeration right-shifts and dilates the distribution. Using this prediction, French establishment level data, and a new quantile approach, we show that firm selection cannot explain spatial productivity differences. This result holds across sectors, city size thresholds, establishment samples, and area definitions.

JEL Classification: C52, R12, D24

Keywords: agglomeration, firm selection, productivity, cities

Corresponding author:

Laurent Gobillon
Institut National d'Etudes Démographiques (INED)
133 Boulevard Davout
75980 Paris Cedex 20
France
E-mail: laurent.gobillon@ined.fr

^{*} Replication files for this paper are available from <http://diegopuga.org/data/selectagg/>. We thank Kristian Behrens, Steven Berry, Stéphane Grégoir, Robert McMillan, Marc Melitz, Peter Neary, Gianmarco Ottaviano, Giovanni Peri, Stephen Redding, John Sutton, Daniel Treffer, five anonymous referees, and conference and seminar participants for comments and discussions. We gratefully acknowledge funding from the Agence Nationale de la Recherche through grant COMPNASTA (Combes), the Banco de España Excellence Programme (Puga), the Canadian Social Science and Humanities Research Council (Duranton), the Centre National de la Recherche Scientifique (Combes), the Comunidad de Madrid through grant S2007/HUM/0448 PROCIUDAD-CM (Duranton and Puga), the European Commission's Seventh Research Framework Programme through contract number 269868 for the European Research Council's Advanced Grant 'Spatial Spikes' (Puga), and through contract number 225551 for the Collaborative Project 'European Firms in a Global Economy (efige)' (Combes, Duranton, Gobillon, and Puga), and the Fundación Ramón Areces (Puga).

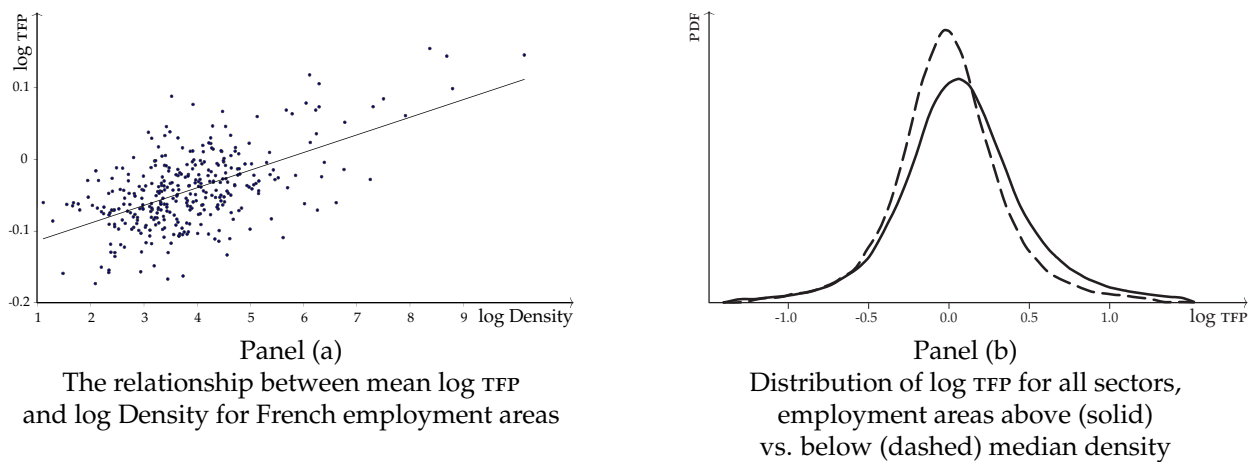


Figure 1: The productive advantages of large cities

1. Introduction

Firms and workers are, on average, more productive in larger cities. This fact — already discussed by Adam Smith (1776) and Alfred Marshall (1890) — is now firmly established empirically (see Rosenthal and Strange, 2004, and Melo, Graham, and Noland, 2009, for reviews and summaries of existing findings). Estimates of the elasticity of productivity with respect to city size range between 0.02 and 0.10, depending on the sector and details of the estimation procedure. Panel (a) of figure 1 illustrates this by plotting mean log TFP against log employment density (workers per square kilometre), the most common measure of local scale in the literature, for all 341 employment areas in continental France in 1994–2002. On this plot, the slope of the regression line is 0.025 and the R^2 is 0.33. Panel (b) of figure 1 shows the distribution of log TFP in employment areas with above-median employment density and below-median employment density. While a full discussion is provided below, one can immediately see that the higher mean log TFP in denser employment areas is accounted for by changes over the entire distribution. Figure 2 maps the geography underlying panel (a) of figure 1, with log employment density shown in panel (a) and mean log TFP in panel (b) of figure 2.

For a long time, the higher average productivity of firms and workers in larger cities has been attributed to ‘agglomeration economies’. These agglomeration economies are thought to arise from a variety of mechanisms, such as the possibility for similar firms to share suppliers, the existence of thick labour markets ironing out firm-level shocks or facilitating matching, or the possibility to learn from the experiences and innovations of others. All these agglomeration mechanisms share a common prediction: the concentration of firms and workers in space makes them more productive (see Duranton and Puga, 2004, for a review). While studying agglomeration mechanisms, urban economists have kept in mind that the productivity advantage of larger cities could also be explained by localised natural advantage or the sorting of more able workers.

More recently, an alternative explanation has been offered. It is based on ‘firm selection’ and builds on work by Melitz (2003), who introduces product differentiation and international or inter-regional trade into the framework of industry dynamics of Hopenhayn (1992). Melitz and Ottaviano (2008) incorporate variable price-cost mark-ups in this framework and show that larger

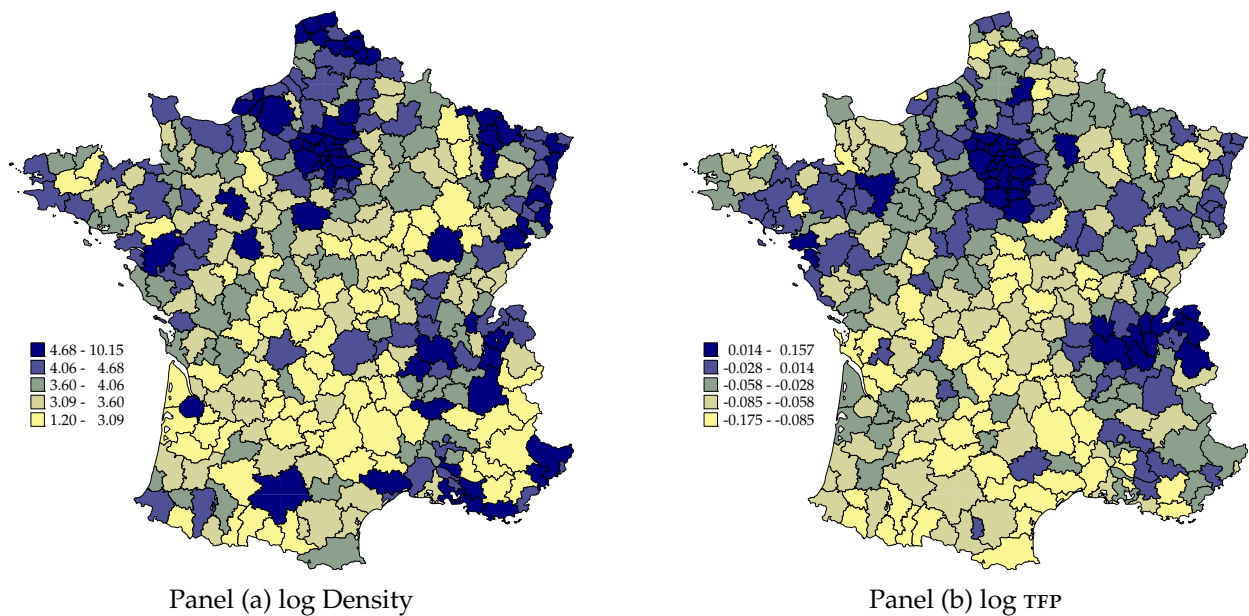


Figure 2: Geographic distribution of log employment density and mean log TFP in France

markets attract more firms, which makes competition tougher. In turn, this leads less productive firms to exit. This suggests that the higher average productivity of firms and workers in larger cities could result instead from a stronger Darwinian selection of firms.

Our main objective in this paper is to distinguish firm selection from the other motives behind the productive advantages of cities commonly considered by urban economists. In our context ‘selection’ is the inability of weak firms to survive when faced with tougher competition in larger markets.¹ Anticipating our results, we find that selection is not important for explaining productivity differences across cities relative to the productivity advantages commonly studied by urban economists. However, we do not tackle which of those common advantages (sharing, matching, learning, localised natural advantages) are more important. We also do not deal here with issues raised by the possible endogeneity of city scale, although evidence from other research indicates such issues, if anything, are minor.² Furthermore, even if localised natural advantages were causing both differences in scale and productivity distributions across cities, it is still interesting not to find more selection in cities with higher average productivity.

The first step of our approach is to free the framework of Melitz and Ottaviano (2008) from distributional assumptions and generalise it to many cities. We then combine this model with a fairly general model of agglomeration in the spirit of Fujita and Ogawa (1982) and Lucas and

¹The term ‘selection’ is sometimes used to refer to different processes. In Nocke (2006), more able entrepreneurs sort into larger markets where competition becomes more intense. In Baldwin and Okubo (2006), it is more productive firms that sort into larger markets because they benefit more from forward and backward linkages. Holmes, Hsu, and Lee (2011) develop the framework of Bernard, Eaton, Jensen, and Kortum (2003). Selection in their model makes mark-ups lower in larger markets but the number of surviving firms remains constant at one per market.

²Extant research attempts to separate agglomeration from localised natural advantage using instrumental variables (following Ciccone and Hall, 1996), panel data (after Henderson, 1997), explicit controls for natural advantages (e.g., Ellison and Glaeser, 1999), or a combination of the above (Combes, Duranton, Gobillon, and Roux, 2010, who use the same data as this paper). All three approaches lead those papers to similar conclusions. While localised natural advantages are possibly important location determinants, accounting for them as carefully as possible does not detract much from the estimated magnitude of agglomeration economies.

Rossi-Hansberg (2002). This nested model allows us to parameterise the relative importance of agglomeration and selection.

While this model makes specific assumptions about market structure, production, trade costs, and demand, our empirical approach builds on two properties that we expect to hold more widely. If selection is tougher in larger cities, fewer of the weaker firms will survive there. Stronger selection should thus lead to a greater left truncation of the distribution of firm log productivity in larger cities. If agglomeration economies are stronger in larger cities, all firms located there will enjoy some productive advantages, with perhaps some benefiting more than others. Stronger agglomeration effects in larger cities should thus lead instead to a greater rightwards shift of the distribution of firm log productivity in larger cities. To the extent that more productive firms are better able to reap the benefits of agglomeration, agglomeration should also lead to an increased dilation of the distribution of firm log productivity in larger cities. While these properties should hold more generally, our structural model helps interpret the empirical results.

We then use these predictions to assess the extent to which selection, as opposed to agglomeration economies or localised natural advantages, drives productivity differences across French employment areas. Our estimation relies on two identification conditions, namely a common underlying productivity distribution for potential entrants and separability between agglomeration and selection. We proceed in two steps. We first estimate total factor productivity at the establishment level. Next, we develop a new quantile approach to compare the distribution of establishment log productivity for each sector across French areas of different density. Panel (b) of figure 1 plots the distribution of log TFP for production establishments in manufacturing and business services in employment areas with above-median employment density (solid line) and in employment areas with below-median employment density (dashed line). Since it is hard to separate truncation, shift and dilation in a purely visual comparison of distributions, our approach estimates the extent to which the log productivity distribution in denser areas is left-truncated (evidence of differences in selection effects) or dilated and right-shifted (evidence of common productivity advantages) compared to the log productivity distribution in less dense areas.

This empirical approach offers a number of benefits. First, it allows both firm selection and agglomeration economies to play a role, instead of focusing on just one or the other. Second, while firmly grounded in a nested model, our approach identifies selection and agglomeration from features that are common to a much broader class of models. Basically, it relies on fiercer competition eliminating the weakest firms. Agglomeration economies, as well as localised natural advantages, raise everyone's productivity, possibly to different extents. Third, we do not rely on particular distributional assumptions of firm productivity nor on a particular moment of the data. Fourth, our approach does not attempt to identify selection by looking for cutoffs in the lower tail of the log productivity distribution, which may be obscured by measurement error, nor by looking for lesser log productivity dispersion in larger cities, which is not a necessary consequence of selection. Instead, it estimates differences in truncation across areas from their entire distributions using the fact that greater truncation raises the density distribution proportionately everywhere to the right of the cutoff. Finally, our approach is agnostic as to the different motives for productivity benefits enjoyed by all firms in large cities including agglomeration economies and exogenous

characteristics, but nonetheless allows these benefits to differ systematically across firms.

Our main finding is that there are no sizeable differences in left truncation between denser and less dense employment areas, indicating that selection does not play a major role in explaining the productive advantages of urban density. Instead, the entire log productivity distribution in denser areas is right-shifted relative to the distribution in less dense areas. Furthermore, more productive establishments are better able to reap the benefits of urban density, which dilates the log productivity distribution. As a result, while the average productivity gain is about 9.7 percent, establishments at the bottom quartile of the log productivity distribution are only 4.8 percent more productive in employment areas with above-median density than elsewhere whereas establishments at the top quartile are about 14.4 percent more productive in denser areas. These results are robust to changes in the choice of estimation technique for productivity, the sample of establishments, the choice of spatial units, and the measure of local scale.

Our paper is related to the pioneering work by Syverson (2004), who examines the effect of market size on firm selection in the ready-made concrete sector, and the emerging literature that follows (e.g., Del Gatto, Ottaviano, and Pagnini, 2008). A first difference with Syverson's work is that we build our empirical approach on a nested model of selection and agglomeration rather than a model incorporating selection alone. Considering more traditional productive advantages of large cities simultaneously with selection allows us to identify robust differences in predictions between the two types of mechanisms. A second difference is that, instead of examining differences in summary statistics across locations, we develop a quantile approach that traces differences throughout the log productivity distribution. A third difference is that we consider firms not only in the ready-made concrete sector but in the entire economy. Our work is also related to the large agglomeration literature building on Henderson (1974) and Sveikauskas (1975), and surveyed in Duranton and Puga (2004), Rosenthal and Strange (2004) and Head and Mayer (2004). We extend it by considering an entirely different reason for the higher average productivity in larger cities. Our paper is finally related to Carrasco and Florens (2000), since our quantile approach adapts their results for an infinite set of moments to deal with an infinite set of quantile equalities.³

The rest of this paper is organised as follows. The next section develops a nested model of selection and agglomeration. Section 3 describes our econometric approach. Section 4 discusses the data and the details of our empirical implementation. The results are then presented in section 5. Section 6 discusses some additional issues, and section 7 concludes.

³There is also a large literature in international trade that explores whether good firms self-select into exporting or learn from it. Early studies (Clerides, Lach, and Tybout, 1998, Bernard and Jensen, 1999) point to the predominance of self-selection by observing that exporting firms have better pre-determined characteristics. More recent work by Lileeva and Trefler (2010) shows that lower US tariffs provided less productive Canadian firms with an opportunity to invest and improve their productivity to export to the US. A similar type of question can be raised regarding the higher productivity of firms in import-competing sectors. Pavcnik (2002) uses trade liberalisation in Chile to provide evidence about both selection (the exit of the least productive firms and factor reallocation towards the more productive firms) and increases in productivity when firms have to compete with importers. Both strands of literature usually identify selection from changes over time either in trade policy or along the firm life-cycle. With city size changing only slowly over time, we need to use instead a cross-sectional approach. The other difference with the trade literature is that we implement a structural model rather than run reduced-form regressions. We defer further discussion of how our results fit with the implications from this trade literature to the concluding section.

2. A nested model of selection and agglomeration

Our aim is to compare the distribution of firm log productivity across cities of different sizes. To build the theoretical foundations of our empirical approach, we nest a generalised version of the firm selection model of Melitz and Ottaviano (2008) and a model of agglomeration economies along the lines of Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002).

Suppose we have I cities and let us denote the population of city i by N_i . An individual consumer's utility is given by

$$U = q^0 + \alpha \int_{k \in \Omega} q^k dk - \frac{1}{2} \gamma \int_{k \in \Omega} (q^k)^2 dk - \frac{1}{2} \eta \left(\int_{k \in \Omega} q^k dk \right)^2, \quad (1)$$

where q^0 denotes the individual's consumption of a homogenous numéraire good, and q^k her consumption of variety k of a set Ω of differentiated products. The three positive demand parameters α , γ , and η are such that a higher α and a lower η increase demand for differentiated products relative to the numéraire, while a higher γ reflects more product differentiation between varieties.

Utility maximisation yields individual inverse demand for differentiated product k as

$$p^k = \alpha - \gamma q^k - \eta \int_{j \in \Omega} q^j dj, \quad (2)$$

where p^k denotes the price of product k . It follows from (2) that differentiated products with too high a price are not consumed. This is because, by (1), the marginal utility for any particular product is bounded. Let $\bar{\Omega}$ denote the set of products with positive consumption levels in equilibrium, ω the measure of $\bar{\Omega}$, and $P \equiv \frac{1}{\omega} \int_{j \in \bar{\Omega}} p^j dj$ the average price faced by the individual consumer for products with positive consumption. Integrating equation (2) over all products in $\bar{\Omega}$, solving for $\int_{j \in \bar{\Omega}} q^j dj$, and substituting this back into equation (2), we can solve for an individual consumer's demand for product k as

$$q^k = \begin{cases} \frac{1}{\gamma + \eta \omega} (\alpha + \frac{\eta}{\gamma} \omega P) - \frac{1}{\gamma} p^k & \text{if } p^k \leq \bar{h} \equiv P + \frac{\gamma(\alpha - P)}{\gamma + \eta \omega}, \\ 0 & \text{if } p^k > \bar{h}. \end{cases} \quad (3)$$

The price threshold, \bar{h} , in equation (3) follows immediately from the restriction $q^k \geq 0$. By the definition of P and equation (2), $P < \alpha$ so that $\bar{h} > P$.

The numéraire good is produced under constant returns to scale using one unit of labour per unit of output. It can be freely traded across cities. This implies that the cost to firms of hiring one unit of labour is always unity.⁴ Differentiated products are produced under monopolistic competition. By incurring a sunk entry cost s , a firm develops a new product that can be manufactured using h units of labour per unit of output. Given that the cost of each unit of labour equals one unit of the numéraire, h is also the marginal cost. The value of h differs across firms. For each of them h is randomly drawn, after the sunk entry cost has been incurred, from a distribution with known probability density function $g(h)$ and cumulative $G(h)$ common to all cities. Firms with a marginal cost higher than the price at which consumer demand becomes zero are unable to cover

⁴The unit cost for labour holds provided there is some production of the numéraire good everywhere. Given the quasi-linear preferences, this requires that income is high enough, which is easy to ensure.

their marginal cost and exit. The set of products that end up being produced in equilibrium is therefore $\bar{\Omega} = \{k \in \Omega \mid h \leq \bar{h}\}$.

Melitz and Ottaviano (2008) derive most of their results under the assumption that $1/h$ follows a Pareto distribution. By contrast, we do not adopt any particular distribution for $g(h)$. For simplicity, we only require $G(\cdot)$ to be differentiable. Appendix A shows that this generality is important since the empirical distribution of $1/h$ is not well approximated by a Pareto. If anything, it is close to a log-normal with a slightly fatter upper tail. As shown below, the core results of Melitz and Ottaviano (2008) are robust to not assuming a specific distribution.

Suppose that markets for differentiated products are segmented and that selling outside the city where a firm is located involves iceberg trade costs so that $\tau (> 1)$ units need to be shipped for one unit to arrive at destination.⁵ While goods are tradable, we assume that firms are immobile. This is a reasonable approximation of what happens in France, the country for which we implement our empirical exercise.⁶ However, even with limited *ex-post* firm mobility, the *ex-ante* mobility of entrepreneurs may be important. We leave this issue aside here for tractability (see Behrens, Duranton, and Robert-Nicoud, 2010, for a recent step in this direction).

Since all differentiated products enter symmetrically into utility, we can index firms by their unit labour requirement h and their city i instead of their specific product. Indexing now also consumers by their location j , re-writing the individual consumer demand of (3) in terms of \bar{h}_j ,

$$q_{ij}(h) = \frac{1}{\gamma + \eta\omega_j}(\alpha + \frac{\eta}{\gamma}\omega_j P_j) - \frac{1}{\gamma}p_{ij}(h) = \frac{1}{\gamma} \left(P_j + \frac{\gamma(\alpha - P_j)}{\gamma + \eta\omega_j} - p_{ij}(h) \right) = \frac{1}{\gamma}[\bar{h}_j - p_{ij}(h)] ,$$

and multiplying this by the mass of consumers in city j , N_j , yields the following expression for the demand faced in city j by an individual firm from city i with unit requirement h :

$$Q_{ij}(h) = N_j q_{ij}(h) = \frac{N_j}{\gamma}[\bar{h}_j - p_{ij}(h)] \quad \text{if } p_{ij}(h) \leq \bar{h}_j , \quad (4)$$

and $Q_{ij}(h) = 0$ if $p_{ij}(h) > \bar{h}_j$. Given that the entry cost is sunk when firms draw their value of h , a firm from city i with unit requirement h selling in city j sets its price there to maximise operational profits in the city given by $\pi_{ij}(h) = [p_{ij}(h) - \tau_{ij}h]Q_{ij}(h)$, where $\tau_{ij} = 1$ if $i = j$ and $\tau_{ij} = \tau$ if $i \neq j$, subject to (4). This yields

$$p_{ij}(h) = \frac{1}{2}(\bar{h}_j + \tau_{ij}h) . \quad (5)$$

Substituting (4) and (5) into the expression for $\pi_{ij}(h)$ we obtain equilibrium operational profits:

$$\pi_{ij}(h) = \frac{N_j}{4\gamma}(\bar{h}_j - \tau_{ij}h)^2 .$$

Entry into the monopolistically competitive industry takes place until *ex-ante* expected profits from all markets are driven to zero. The operational profits expected prior to entry must therefore be

⁵We assume implicitly that all cities have equal access other cities. Our main theoretical result readily generalises to situations where larger cities have better access to other cities. We also show below that our empirical results are not affected by conditioning out market access.

⁶Duranton and Puga (2001) report that only 4.7% of French establishments change their location to a different employment area over the four years from 1993 to 1996. These moves also appear to be primarily related to firm life-cycle considerations where mature firms move away from large diverse areas to save costs.

exactly offset by the sunk entry cost:

$$\frac{N_i}{4\gamma} \int_0^{\bar{h}_i} (\bar{h}_i - h)^2 g(h) dh + \sum_{j \neq i} \frac{N_j}{4\gamma} \int_0^{\bar{h}_j/\tau} (\bar{h}_j - \tau h)^2 g(h) dh = s, \quad (6)$$

for city i . The first term on the left-hand side captures operational profits from local sales and the second-term summation the operational profits from out-of-city sales. Note that all city i firms with marginal costs $h < \bar{h}_i$ sell locally but only those with $h < \bar{h}_j/\tau$ sell in city j , where \bar{h}_j is the cutoff for local firms in j , since city i firms must be able to cover not just production but also trade costs. Expression (6) provides I free entry equations that implicitly define the I marginal cost cutoffs $\bar{h}_1, \dots, \bar{h}_I$ as a function of city sizes N_1, \dots, N_I , the marginal cost distribution $g(h)$, the sunk entry cost s , and the degree of product differentiation parameter γ .

We now turn to the agglomeration components of the model. Workers are endowed with a single unit of working time each that they supply inelastically. Each worker is made more productive by interactions with other workers. We can think of such interactions as exchanges of ideas between workers, where being exposed to a greater diversity of ideas makes each worker more productive. This motivation for agglomeration economies based on interactions between workers can be found in, amongst others, Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002). As in these papers, interactions are subject to a spatial decay. This implies that the effective labour supplied by an individual worker in city i is $a(N_i + \delta \sum_{j \neq i} N_j)$, where $a(0) = 1$, $a' > 0$, and $a'' < 0$. The decay parameter δ measures the strength of across-city relative to within-city interactions ($0 \leq \delta \leq 1$). This, given the unit payment per effective unit of labour supplied, implies that the total labour income of each worker in any occupation is $a(N_i + \delta \sum_{j \neq i} N_j)$.

A firm in city i with unit labour requirement h hires $l_i(h) = \sum_j Q_{ij}(h)h/a(N_i + \delta \sum_{j \neq i} N_j)$ workers at a total cost of $a(N_i + \delta \sum_{j \neq i} N_j)l(h) = \sum_j Q_{ij}(h)h$. Let

$$A_i \equiv \ln [a(N_i + \delta \sum_{j \neq i} N_j)]. \quad (7)$$

The natural logarithm of the firm's productivity is then given by

$$\phi_i(h) = \ln \left(\frac{\sum_j Q_{ij}(h)}{l_i(h)} \right) = A_i - \ln(h). \quad (8)$$

Let us denote the proportion of firms that fail to survive product-market competition in city i (a local measure of the strength of selection) by

$$S_i \equiv 1 - G(\bar{h}_i). \quad (9)$$

To further simplify notation, let us define $\tilde{F}(\phi) \equiv 1 - G(e^{-\phi})$ as the underlying cumulative density function of log productivity we would observe in all cities in the absence of any selection ($\bar{h}_i \rightarrow \infty, \forall i$) and in the absence of any agglomeration ($A_i = 0, \forall i$). Without selection all entrants survive regardless of their draw of h . Without agglomeration, $\phi = -\ln(h)$. Equivalently, $h = e^{-\phi}$. Using the change of variables theorem then yields $\tilde{F}(\phi)$. We can then write the cumulative density function of the distribution of log productivity for active firms in city i as

$$F_i(\phi) = \max \left\{ 0, \frac{\tilde{F}(\phi - A_i) - S_i}{1 - S_i} \right\}. \quad (10)$$

This follows from equations (8) and (9) and the change of variables theorem. Relative to the underlying distribution given by $\tilde{F}(\phi)$, agglomeration shifts the distribution of log productivity rightwards by A_i while selection eliminates a share S_i of entrants (those with lower productivity values).⁷ The model can now be solved sequentially by first using the free entry conditions of equation (6) to solve for the equilibrium cut-off unit labour requirements \bar{h}_i , for $i = 1, \dots, I$. We can then substitute \bar{h}_i into (9) and S_i into (10) to obtain the equilibrium distribution of firm productivity. Finally, equation (5) gives prices, and the definition of \bar{h}_i in (3) tells us what products are sold in each city, ω_i .

While we treat city sizes as exogenous, these can be endogenised. In a separate web appendix, we show how one can introduce worker mobility, urban crowding costs, and consumption amenities in the spirit of Henderson (1974) and Roback (1982). This provides an additional set of equations relating city sizes to amenities, real wages, and crowding costs, which can be treated independently from the rest of our framework. In this extended model, cities with greater amenities are larger in size and this larger size goes together with higher nominal earnings for workers due to stronger agglomeration economies. On the other hand, larger cities have the disadvantage of higher costs associated with housing and commuting, and in equilibrium city sizes adjust so that the net advantages and disadvantages of larger cities exactly balance out against the value of the amenities they provide. Our main theoretical proposition below holds unchanged in this extended version of the model (see the separate web appendix for details).

Panels (a) and (b) of figure 3 illustrate the model by plotting the distributions of firm log productivity in a city with a large population (continuous line) and in a city with a small population (dashed line) in two polar cases. In panel (a), $\tau \rightarrow \infty$ and $\delta = 1$, so that firms only sell in their local city and workers enjoy interactions with the same intensity with workers from everywhere. As we can see, the large-city distribution is left-truncated relative to the small city distribution. This occurs as a consequence of tougher firm selection. If the number of active firms in the large city was the same as in the small city, every large-city firm would sell proportionately more. However, the larger individual firm sales associated with a larger local market make further entry profitable and, by equation (6), they must be offset by a lower \bar{h} to restore zero *ex-ante* expected profits, leading to left-truncation.⁸ In panel (b), $\tau = 1$ and $\delta = 0$, so that every firm competes with the same intensity with firms from everywhere and workers only interact with workers in their city. As we can see, the large-city distribution is right-shifted relative to the small city distribution. This occurs because

⁷Note that localised natural advantage can have an effect similar to that of stronger agglomeration economies if their magnitude is positively correlated with city size. In particular, we could rewrite equation (7) as $A_i \equiv \ln[\kappa_i a(N_i + \delta \sum_{j \neq i} N_j)]$, where κ_i is a city-specific shifter capturing any kind of localised advantages. A higher κ_i and a higher N_i would then both result in a higher A_i . For this reason, as noted above, we make no strong claim in this paper about the relative importance of externalities and natural advantage. See the introduction for references to the existing empirical literature on this distinction.

⁸By (3), even if firms were to keep their prices constant following entry (leaving P unchanged), the business stealing effect of entry (larger ω) is enough to make the sales of more expensive products drop to zero. In turn, by (5), this lower \bar{h}_i induces firms to lower their prices which, by (3), further reduces \bar{h}_i . Some low-productivity firms that would have been able to survive in a small city cannot lower their prices any further and must exit in the large city.

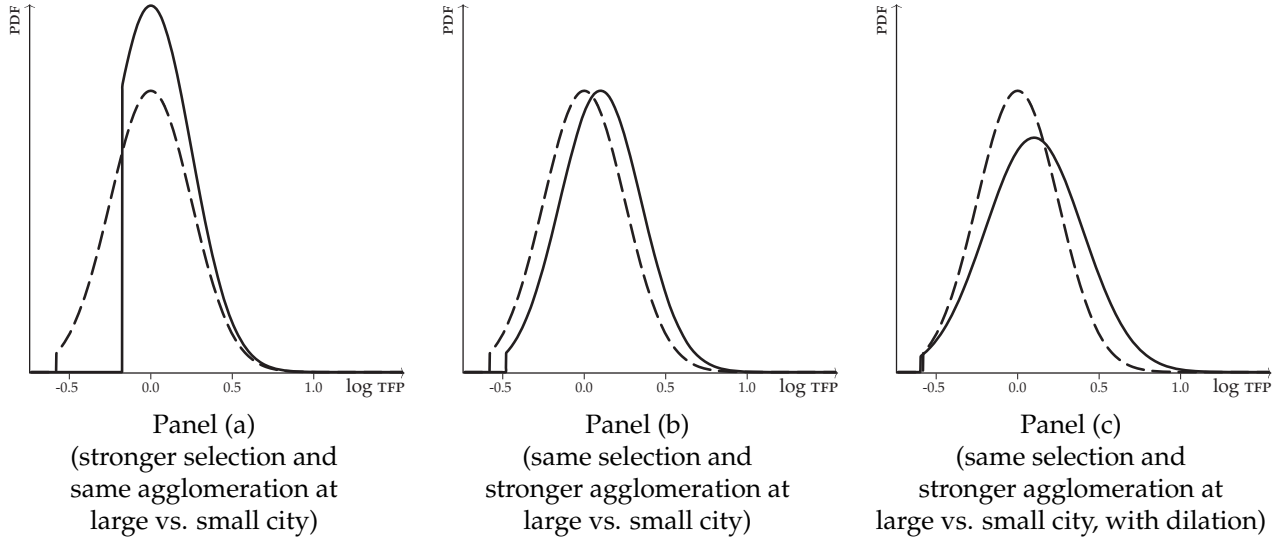


Figure 3: Log TFP distributions in large (solid) and small cities (dashed)

all firms in the large city enjoy the benefits from locating there.⁹

These two examples illustrate a more general proposition with $0 \leq \delta \leq 1$, and $1 \leq \tau < \infty$. However, we would like to derive this proposition in an even more general setting that also allows the magnitude of agglomeration economies to be systematically related to individual productivity and not just to city size. In particular, we conjecture that, while agglomeration economies raise the productivity of all firms in larger cities, the gain is greater for the most productive firms. To capture this idea in a simple way, let us thus relax the assumption that workers are equally productive regardless of the firm they work for. Suppose instead that workers are more productive when they work for a more efficient firm (i.e., one with a lower h) and that this effect is enhanced by interactions with other workers. In particular suppose that the effective units of labour supplied by an individual worker in their unit working time are $a(N_i + \delta \sum_{j \neq i} N_j)h^{-(D_i-1)}$, where

$$D_i \equiv \ln [d(N_i + \delta \sum_{j \neq i} N_j)] , \quad (11)$$

$d(0) = 1$, $d' > 0$ and $d'' < 0$ (the model seen up until this point was equivalent to assuming $D_i = 1$). In this case, the natural logarithm of the productivity of a firm with unit cost h in city i is

$$\phi_i(h) = \ln \left(\frac{\sum_j Q_{ij}(h)}{l_i(h)} \right) = A_i - D_i \ln(h) . \quad (12)$$

We can then write the cumulative density function of the distribution of log productivity for active firms in city i as

$$F_i(\phi) = \max \left\{ 0, \frac{\tilde{F} \left(\frac{\phi - A_i}{D_i} \right) - S_i}{1 - S_i} \right\} . \quad (13)$$

⁹The examples of figure 3 are simulations of the model, using a log-normal distribution for $g(h)$ (which appendix A shows is a good empirical approximation). In panel (a), $A_i = A_j = 0$ (no differences in agglomeration), $S_i = 0.24$ and $S_j = 0.01$ (which correspond to the differences in selection required to obtain the differences in mean log TFP between denser and less dense cities observed in our data). In panel (b), $S_i = S_j = 0.01$ (no differences in selection), $A_i = 0.1$ and $A_j = 0$ (which correspond to the differences in agglomeration required to obtain the empirical differences in mean log TFP).

Relative to the underlying log productivity distribution \tilde{F} , agglomeration both dilates the distribution by a factor D_i and shifts it rightwards by A_i , while selection eliminates a share S_i of entrants (those with lower productivity values). Panel (c) of figure 3 plots the example of panel (b), with the same selection but stronger agglomeration in the large city, once we allow more productive firms to benefit more from larger cities. The distribution of log productivity in large cities is now both right-shifted and dilated relative to the distribution in small cities as a result of agglomeration economies.¹⁰

The following proposition contains our main theoretical result.

Proposition 1. Suppose there are I cities ranked from largest to smallest in terms of population: $N_1 > N_2 > \dots > N_{I-1} > N_I$, that workers are more productive when they work for a more efficient (lower h) firm and that this effect is enhanced by interactions, that interactions across cities decay by a factor δ , where $0 \leq \delta \leq 1$, and that selling in a different city raises variable costs by a factor τ , where $1 \leq \tau < \infty$.

- i. Agglomeration leads to the distribution of log productivity being dilated by a factor D_i and right-shifted by A_i , and if $\delta < 1$ this dilation and right shift are both greater the larger a city's population: $D_1 > D_2 > \dots > D_{I-1} > D_I$ and $A_1 > A_2 > \dots > A_{I-1} > A_I$.
- ii. Firm selection left-truncates a share S_i of the distribution of log productivity, and if $\tau > 1$ this truncation is greater the larger a city's population: $S_1 > S_2 > \dots > S_{I-1} > S_I$.
- iii. If there is no decay in interactions across cities, so that $\delta = 1$, then there are no differences in dilation nor in shift across cities: $D_i = D_j$ and $A_i = A_j, \forall i, j$. If there is no additional cost incurred when selling in a different city, so that $\tau = 1$, then there are no differences in truncation across cities: $S_i = S_j, \forall i, j$.

Proof See appendix B. □

While this model makes specific assumptions about market structure, production, trade costs, and demand, our empirical approach builds on two properties that we expect to hold more widely. If selection is tougher in larger cities, fewer of the weaker firms will survive there. If agglomeration economies are stronger in larger cities, all firms located there will enjoy some productive advantages, with perhaps some benefiting more than others. The empirical approach we develop next exploits these two properties. It relies on two identification conditions, namely

¹⁰The simulation in panel (c) uses parameters consistent with our preferred empirical estimates, rounded to the first decimal: $A_i = 0.1$, $A_j = 0$, $D_i = 1.2$, $D_j = 1$, and $S_i = S_j = 0.01$. Trying to identify left-truncation, shift, and dilation by visual comparison of two distributions is difficult, and this is why we develop our empirical methodology. However, comparing the peaks of the distributions provides useful visual cues. Left-truncation obviously takes probability mass away from the left of the distribution. In addition, since the area under the probability distribution function still has to integrate to one, this raises the curve proportionately everywhere to the right of the truncation point, and the peak of the distribution (provided the truncation point is below the median) is higher but vertically above the peak of the distribution prior to left-truncation, as in panel (a) of figure 3. A right-shift, on the other hand, moves the entire distribution rightwards leaving the peak at the same height but further to the right, as in panel (b) of figure 3. Dilation, by stretching out the distribution, brings the peak lower (as opposed to truncation, which makes it higher).

a common underlying productivity distribution for potential entrants and separability between agglomeration and selection.¹¹

3. Econometric approach

We now develop an econometric approach to estimate what combination of shift, dilation and left-truncation best explains differences in the distribution of log productivity across cities of different sizes. Ideally, we would like to use the cumulative distribution of log productivity to estimate parameters A_i , D_i , and S_i from equation (13) for each city. However, this is not possible because the baseline cumulative of log productivity \tilde{F} is not observed. Nevertheless, the following lemma shows that we can get around this issue by comparing the distribution of log productivity across two cities of different sizes i and j to difference out \tilde{F} from equation (13).¹²

Lemma 1. Consider two distributions with cumulative density functions F_i and F_j . Suppose F_i can be obtained by dilating by a factor D_i and shifting rightwards by A_i some underlying distribution with cumulative density function \tilde{F} and also left-truncating a share $S_i \in [0,1)$ of its values, as described by equation (13). Suppose F_j can be obtained by dilating by a factor D_j and shifting rightwards by a value A_j the same underlying distribution \tilde{F} and also left-truncating a share $S_j \in [0,1)$ of its values, as would be described by equation (13) after replacing subindex i with j . Let

$$D \equiv \frac{D_i}{D_j}, \quad A \equiv A_i - DA_j, \quad S \equiv \frac{S_i - S_j}{1 - S_j}.$$

If $S_i > S_j$, then F_i can also be obtained by dilating F_j by D , shifting it by A , and left-truncating a share S of its values:

$$F_i(\phi) = \max \left\{ 0, \frac{F_j\left(\frac{\phi - A}{D}\right) - S}{1 - S} \right\}. \quad (14)$$

¹¹As shown below, our empirical finding of no differences in truncation makes separability less crucial. However, beyond our particular application, with significant differences in selection, interactions between agglomeration and selection complicate the evaluation of their exact relative magnitudes. For example, in our model, the absence of interactions between selection and agglomeration mechanisms is a consequence of having kept the assumption of quasi-linear preferences of Melitz and Ottaviano (2008). This eliminates income effects in the market for differentiated products. The introduction of income effects would create an interaction between agglomeration and firm selection that would result in further left truncation of the large-city log productivity distribution. This is because, with income effects, the productivity advantages of agglomeration would translate into a larger market for differentiated products in the large city. This would reinforce the increase in local product-market competition caused by the larger population, and strengthen firm selection. Thus, with income effects, agglomeration would appear as a right shift in the log productivity distribution, while selection as well as interactions between selection and agglomeration would appear as a left truncation.

¹²In the model above, firms that draw too high a value of unit costs never begin production. In practice, firms may not realise what their actual costs are until they have been producing for at least a short period. This suggests that studying differences in early exit rates across areas might provide some information about the relative importance of market selection. However, high exit rates in larger cities could also be the outcome of the following alternative explanation. Large diverse metropolitan areas facilitate learning and experimentation at the early stages of a firm's life cycle, while small specialised areas save costs at more mature stages. This alternative explanation predicts not just higher exit rates in larger cities, but also higher entry rates in larger cities, and a pattern of relocation over a firm's life cycle from larger diverse metropolitan areas to smaller more specialised cities. See Duranton and Puga (2001) for a dynamic urban model where this mechanism operates as well as for evidence that its two additional predictions hold empirically.

If $S_i < S_j$, then F_j can also be obtained by dilating F_i by $\frac{1}{D}$, shifting it rightwards by $-\frac{A}{D}$ and left-truncating a share $\frac{-S}{1-S}$ of its values:

$$F_j(\phi) = \max \left\{ 0, \frac{F_i(D\phi + A) - \frac{-S}{1-S}}{1 - \frac{-S}{1-S}} \right\}. \quad (15)$$

Proof See appendix C. □

We are going to use (14) and (15) to get an econometric specification that can be estimated from the data. An advantage of our approach is that we do not need to specify an ad-hoc underlying distribution of log productivity \tilde{F} , which one cannot observe empirically. A limitation is that we are not able to separately identify A_i , A_j , D_i , D_j , S_i and S_j from the data, but only $A = A_i - DA_j$, $D = D_i/D_j$, and $S = (S_i - S_j)/(1 - S_j)$. In other words, we are able to make statements about the relative strength of firm selection in large cities compared to small cities, but not about its absolute strength. Parameter A measures how much stronger is the right shift in city i relative to the smaller city j . Note that our empirical approach also allows for the possibility that $A < 0$, in which case there would be less rather than more right shift in larger cities. Parameter D measures the ratio of dilation in city i relative to the smaller city j . Again, our empirical approach allows for the possibility that $D < 1$. Parameter S measures how much stronger is the left truncation in city i relative to the smaller city j . In particular, it corresponds to the difference between cities i and j in the share of entrants eliminated by selection, relative to share of surviving entrants in city j . Note that our empirical approach also allows for the possibility that $S < 0$, in which case there would be less rather than more left truncation in larger cities.

A quantile specification

To obtain the key relationship to be estimated, we rewrite the two equations (14) and (15) in quantiles and combine them into a single expression. Assuming that \tilde{F} is invertible, F_i and F_j are also invertible. We can then introduce $\lambda_i(u) \equiv F_i^{-1}(u)$ to denote the u^{th} quantile of F_i and $\lambda_j(u) \equiv F_j^{-1}(u)$ to denote the u^{th} quantile of F_j . If $S > 0$, equation (14) applies and can be rewritten

$$\lambda_i(u) = D\lambda_j(S + (1 - S)u) + A, \quad \text{for } u \in [0, 1]. \quad (16)$$

If $S < 0$, equation (15) applies and can be rewritten

$$\lambda_j(u) = \frac{1}{D}\lambda_i\left(\frac{u - S}{1 - S}\right) - \frac{A}{D}, \quad \text{for } u \in [0, 1]. \quad (17)$$

Making the change of variable $u \rightarrow S + (1 - S)u$ in (17), this becomes

$$\lambda_j(S + (1 - S)u) = \frac{1}{D}\lambda_i(u) - \frac{A}{D}, \quad \text{for } u \in \left[\frac{-S}{1 - S}, 1\right]. \quad (18)$$

We can then write the following equation that combines (16) and (18):

$$\lambda_i(u) = D\lambda_j(S + (1 - S)u) + A, \quad \text{for } u \in \left[\max\left(0, \frac{-S}{1 - S}\right), 1\right]. \quad (19)$$

Equation (19) cannot be directly used for the estimation because the set of ranks $[\max(0, \frac{-S}{1-S}), 1]$ depends on the true value of S , which is not known. We thus make a final change of variable $u \rightarrow r_S(u)$, where $r_S(u) = \max(0, \frac{-S}{1-S}) + [1 - \max(0, \frac{-S}{1-S})] u$, which transforms (19) into

$$\lambda_i(r_S(u)) = D\lambda_j(S + (1-S)r_S(u)) + A, \quad \text{for } u \in [0, 1]. \quad (20)$$

Equation (20) provides the key relationship that we wish to fit to the data. It states how the quantiles of the log productivity distribution in a large city i are related to the quantiles of the log productivity distribution in a small city j via the relative shift parameter A , the relative dilation parameter D , and the relative truncation parameter S .

A suitable class of estimators

To estimate A , D , and S , we use the infinite set of equalities given by (20) which can be rewritten in more general terms as $m_\theta(u) = 0$ for $u \in [0, 1]$, where $\theta = (A, D, S)$ and

$$m_\theta(u) = \lambda_i(r_S(u)) - D\lambda_j(S + (1-S)r_S(u)) - A. \quad (21)$$

We turn to a class of estimators studied by Gobillon and Roux (2010) who adapt to an infinite set of equalities the results derived by Carrasco and Florens (2000) for an infinite set of moments. Let $\hat{m}_\theta(u)$ denote the empirical counterpart of $m_\theta(u)$, where the true quantiles λ_i and λ_j have been replaced by some estimators $\hat{\lambda}_i$ and $\hat{\lambda}_j$ (see the separate web appendix for details on how these estimators are constructed). We can then introduce an error minimisation criterium based on a quadratic norm of functions, following Carrasco and Florens (2000). Let \mathcal{L}^2 denote the set of $[0, 1]^2$ integrable functions, $\langle \cdot, \cdot \rangle$ denote the inner product such that for any functions y and z in \mathcal{L}^2 , we have: $\langle y, z \rangle = \int_0^1 \int_0^1 y(u)z(v)du dv$, and $\| \cdot \|$ denote the corresponding norm. Consider a linear bounded operator B on \mathcal{L}^2 . Let B^* denote its self-adjoint, such that we have: $\langle By, z \rangle = \langle y, B^*z \rangle$. Then, B^*B can be defined through a weighting function $\ell(\cdot, \cdot)$ such that: $(B^*By)(v) = \int_0^1 y(u)\ell(v, u)du$ and thus $\|By\|^2 = \int_0^1 \int_0^1 y(u)\ell(v, u)y(v)du dv$. Let $n = (n_i, n_j)'$, where n_i and n_j denote respectively the number of observations of F_i and F_j . The vector of parameters θ can then be estimated as

$$\hat{\theta} = \arg \min_{\theta} \|B_n \hat{m}_\theta\|,$$

where B_n is a sequence of bounded linear operators.¹³ In the separate web appendix, we show that the vector of estimated parameters $\hat{\theta}$ is consistent and asymptotically normal under standard regularity assumptions.

Implementation

The weights $\ell(v, u)$ leading to the optimal estimator cannot be used in practice because they depend on the true value of the parameters θ . Alternatively, one can rely on a simple weighting

¹³The following mild assumption is made to ensure that the model described by $m_\theta(u) = 0$ for $u \in [0, 1]$ is identified: there exist K ranks (as many as parameters we wish to estimate) u_i, \dots, u_K such that the system $m_\theta(u_i) = 0$ for $i = 1, \dots, K$ admits a unique solution in θ .

scheme such that $\ell(v, u) = 0$ for $u \neq v$ and $\ell(v, v) = \delta_d$ where δ_d is a Dirac mass. With this weighting scheme, the estimator simplifies to:

$$\hat{\theta} = \arg \min_{\theta} \left(\int_0^1 [\hat{m}_{\theta}(u)]^2 du \right) .$$

This estimator is the mean-square error on m_{θ} . However, it has the undesirable feature that it treats the quantiles of the two distributions asymmetrically. In particular, it compares the quantiles of the actual city i log productivity distribution to the quantiles of a left-truncated and right-shifted city j distribution, when it would also be possible to compare the quantiles of the actual city j distribution to the quantiles of a modified city i distribution. We thus implement a more robust estimation procedure that treats the quantiles of the two distributions symmetrically. As a first step, we derive an alternative set of equations to (20) for this reverse comparison. Making the change of variable $u \rightarrow \frac{u-S}{1-S}$ in (16), this becomes

$$\lambda_j(u) = \frac{1}{D} \lambda_i \left(\frac{u-S}{1-S} \right) - \frac{A}{D} , \quad \text{for } u \in [S, 1] . \quad (22)$$

We can then write the following alternative equation to (19) that combines (17) and (22):

$$\lambda_j(u) = \frac{1}{D} \lambda_i \left(\frac{u-S}{1-S} \right) - \frac{A}{D} , \quad \text{for } u \in [\max(0, S), 1] . \quad (23)$$

Let $\tilde{r}_S(u) = \max(0, S) + [1 - \max(0, S)]u$. With a final change of variable $u \rightarrow \tilde{r}_S(u)$ on (23), this provides a new set of equalities $\tilde{m}_{\theta}(u) = 0$, for $u \in [0, 1]$, where

$$\tilde{m}_{\theta}(u) = \lambda_j(\tilde{r}_S(u)) - \frac{1}{D} \lambda_i \left(\frac{\tilde{r}_S(u) - S}{1-S} \right) + \frac{A}{D} . \quad (24)$$

Let $\hat{m}_{\theta}(u)$ denote the empirical counterpart of $\tilde{m}_{\theta}(u)$, where the true quantiles λ_i and λ_j have been replaced by some estimators $\hat{\lambda}_i$ and $\hat{\lambda}_j$. The estimator we actually use is then

$$\hat{\theta} = \arg \min_{\theta} M(\theta) , \quad \text{where} \quad M(\theta) = \int_0^1 [\hat{m}_{\theta}(u)]^2 du + \int_0^1 [\hat{m}_{\theta}(u)]^2 du . \quad (25)$$

In the results below, we report $\hat{\theta} = (\hat{A}, \hat{D}, \hat{S})$ and a measure of goodness of fit $R^2 = 1 - \frac{M(\hat{A}, \hat{D}, \hat{S})}{M(0, 1, 0)}$. This measures what share of the mean squared quantile differences between the large and small city distributions is accounted for by \hat{A} , \hat{D} , and \hat{S} . Standard errors of the estimated parameters are bootstrapped drawing observations for some establishments out of the log productivity distribution with replacement. For each bootstrap iteration, we first re-estimate TFP for each observation employed in the iteration, and we then re-estimate θ . Finally, we use the distribution of estimates of θ that results from all bootstrap iterations to compute the standard errors.

4. Data and TFP estimation

To construct our data for 1994–2002, we merge together three large-scale, French, administrative data sets from the French national statistical institute (INSEE).

The first is BRN-RSI ('Bénéfices Réels Normaux' and 'Régime simplifié d'imposition') which contains annual information on the balance sheet of all French firms, declared for tax purposes. We extract information about each firm's output and use of intermediate goods and materials to compute a reliable measure of value added for each firm and year. We also retain information about the value of all assets to compute a measure of capital, using the reported book values at historical costs. The sector of activity at the three-digit level is also available and a unique identifier for each firm serves to match these data with the other two data sets.

The second data set is SIREN ('Système d'Identification du Répertoire des ENtreprises') which contains annual information on all French private sector establishments, excluding finance and insurance. From this data set, we retain the establishment identifier, the identifier of its firm (for matching with BRN-RSI), and the municipality where the establishment is located.

The third data set is DADS ('Déclarations Annuelles de Données Sociales'), a matched employer-employee data set, which is exhaustive during the study period. This includes the number of paid hours for each employee in each establishment and her two digit occupational category, which allows us to take labour quality into account. The procedure of Burnod and Chenu (2001) is then used to aggregate total hours worked at each establishment by workers in each of three skill groups: high, intermediate and low skills. The separate web appendix contains further details.

To sum up, for each firm and each year between 1994 and 2002, we know the firm's value added, the value of its capital, and its sector of activity. For each establishment within each firm, we know its location, and the number of hours worked by its employees by skill level. We retain information on all establishments from all firms with 6 employees or more in all manufacturing sectors and in business services, with the exception of finance and insurance (for which individual establishment data is not available).¹⁴ We end up with data on 148,705 firms and 166,086 establishments observed at least once during the study period. See the separate web appendix for further details about the data.

We implement our approach on two different sets of French geographical units: employment areas and urban areas. The 341 French employment areas entirely cover continental France and might be taken as a good approximation for local labour markets. The 364 French urban areas only cover part of continental France and correspond to metropolitan areas. To capture urban scale, population size is natural for urban areas but employment density is more natural for employment areas which sometimes comprise only a part of a metropolitan area. Since we prefer to cover the entire country, for our baseline estimates we lump employment areas together based on their employment density and we compare the distribution of firm log productivity in employment areas with above-median density with the corresponding distribution in employment areas with below-median employment density. We then check the robustness of our results to finer groupings

¹⁴Whenever one estimates firm-level TFP, measurement errors are likely to result in a few extreme outliers. To minimise the impact of such outliers in our estimates of A , D , and S , we exclude the 1 percent of observations with the highest TFP values and the 1 percent of observations with the lowest TFP values in each city size class. It is important to trim extreme values in both city size classes to avoid biasing the estimate of S . Thus, we end up with 162,765 establishments (98 percent of 166,086) in the estimations that combine all establishments from all sectors (bottom panel of table III). Noisier estimates are also the reason for not including establishments with one to five employees in our baseline estimations, although we report estimates for them in table VI. We discuss in detail the implications of noisy TFP estimates for our methodology in section 6.

of employment areas, to the use of urban areas instead of employment areas as spatial units, and to the use of population size instead of employment density as our criterion for grouping spatial units.

TFP estimation

For simplicity of exposition, we have set up the model of section 2 so that labour is the only input. However, all results extend trivially to a model with capital and workers with multiple skill levels, provided technology is homothetic, capital costs are equal at all locations, and from the point of view of an individual firm multiple types of workers are perfect substitutes (up to a scaling factor to capture the impact of skills on efficiency units).

For the purpose of estimation, we assume more specifically that the technology to generate value added at the firm level (V_t) is Cobb-Douglas in the firm's capital (k_t) and labour (l_t), and use t to index time (years). We also allow for three skill levels, and use $l_{s,t}$ to denote the share of the firm's workers with skilled level s :

$$V_t = (k_t)^{\beta_1} \left(l_t \sum_{s=1}^3 \varsigma_s l_{s,t} \right)^{\beta_2} e^{\beta_{3,t} + \phi_t} ,$$

where β_1 , β_2 and the three ς_s are common to all firms within a sector, $\beta_{3,t}$ varies by detailed subsector of that sector, and ϕ_t is firm-specific. Taking logs yields

$$\ln(V_t) = \beta_1 \ln(k_t) + \beta_2 \ln(l_t) + \beta_2 \ln \left(\sum_{s=1}^3 \varsigma_s l_{s,t} \right) + \beta_{3,t} + \phi_t . \quad (26)$$

To linearise (26), we use the approximation in Hellerstein, Neumark, and Troske (1999). If the share of labour with each skill does not vary much over time or across firms within each sector, so that $l_{s,t} \approx \tilde{\varsigma}_s$, then

$$\beta_2 \ln \left(\sum_{s=1}^3 \varsigma_s l_{s,t} \right) \approx \beta_2 \left[\ln \left(\sum_{s=1}^3 \varsigma_s \tilde{\varsigma}_s \right) - 1 \right] + \sum_{s=1}^3 \sigma_s l_{s,t} , \quad (27)$$

where $\sigma_s \equiv \beta_2 \varsigma_s / (\sum_{s=1}^3 \varsigma_s \tilde{\varsigma}_s)$. Substituting equation (27) into (26) yields:

$$\ln(V_t) = \beta_{0,t} + \beta_1 \ln(k_t) + \beta_2 \ln(l_t) + \sum_{s=1}^3 \sigma_s l_{s,t} + \phi_t , \quad (28)$$

where $\beta_{0,t} \equiv \beta_{3,t} + \beta_2 [\ln(\sum_{s=1}^3 \varsigma_s \tilde{\varsigma}_s) - 1]$.

We obtain log TFP by estimating equation (28) separately for each sector in level 2 of the Nomenclature Economique de Synthèse (NES) sectoral classification, which leaves us with 16 manufacturing sectors and business services. We let $\beta_{0,t}$ be the sum of a year-specific component and a sector-specific component at level 3 of the NES classification (which contains 63 subsectors for our base 16 sectors). Denote by $\hat{\beta}_{0,t}$, $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\sigma}_s$ the estimates of $\beta_{0,t}$, β_1 , β_2 and σ_s , respectively. Let $\hat{\phi}_t = \ln(V_t) - \hat{\beta}_{0,t} - \hat{\beta}_1 \ln(k_t) - \hat{\beta}_2 \ln(l_t) - \sum_{s=1}^3 \hat{\sigma}_s l_{s,t}$. We then measure log TFP for each firm by the firm-level average of $\hat{\phi}_t$ over the period 1994–2002,

$$\hat{\phi} = \frac{1}{T} \sum_{t=1}^T \hat{\phi}_t ,$$

where T denotes the number of years the firm is observed in 1994–2002.

For our baseline results, we estimate equation (28) using ordinary least squares (OLS). Later, we report as robustness checks the results obtained with the methods proposed by Olley and Pakes (1996) and Levinsohn and Petrin (2003) to account for the potential endogeneity of capital and labour, as well as simple cost share estimates of TFP. Details on how TFP estimates are constructed in our context using these methods are relegated to a separate web appendix. While each of the different methods to estimate TFP has its own advantages and potential problems, we note our results are completely robust to using any of the established methods in the literature.

Since data for value added and capital is only available at the firm level, in the baseline results we restrict the sample to firms with a single establishment (which account for 92 percent of firms, 82 percent of establishments, and 54 percent of average employment over the period). Later, we report as robustness checks results for all firms, including those with establishments in multiple locations. We do so by estimating the following relationship between each firm's log TFP and the set of locations where it has establishments, separately for each sector:

$$\hat{\phi} = \sum_{i=1}^I \nu_i l_i + \epsilon ,$$

where i indexes locations, and l_i denotes the share of a firm's labour (in hours worked) in location i , averaged over the period 1994–2002. Parameter ν_i is common to all firms and establishments in location i . Let $\hat{\nu}_i$ be the OLS estimate of ν_i and $\hat{\epsilon} = \hat{\phi} - \sum_{i=1}^I \hat{\nu}_i l_i$. Establishment-level log TFP is then computed as $\hat{\nu}_i + \hat{\epsilon}$. Note that for firms with a single establishment, $\hat{\nu}_i + \hat{\epsilon} = \hat{\phi}$ as before.

5. Results

Our main results are presented in the first part of this section. They report estimates of how the distribution of firm log productivity in employment areas with above-median density is best approximated by shifting, dilating and truncating the distribution of firm log productivity in employment areas with below-median density. These results are for 16 manufacturing and business service sectors and for all sectors together using OLS TFP estimates for mono-establishment firms. In the second part of this section, we consider a number of robustness checks.

Columns (1), (2), and (3) of table 1 report our estimates of A , D , and S together with bootstrapped standard errors. The value of A corresponds to the average increase in log productivity that would arise in denser relative to less dense employment areas absent any selection.¹⁵ When $A > 0$, values of D above unity are evidence that the more productive firms benefit more from being in denser employment areas. Values of D below unity would indicate that the more productive firms benefit less from being in denser employment areas. Positive values of S correspond to the distribution of firm log productivity in denser employment areas being more truncated than in less dense employment areas. Negative values correspond instead to more truncation in less dense employment areas.

¹⁵We normalise our log-TFP estimates so that our estimates of A can be interpreted as the average increase in productivity enjoyed by firms in denser employment areas relative to less dense employment areas. This involves choosing units of value added so that average log-TFP in less dense employment areas is zero, which affects neither D nor S .

Table 1: Main estimation results, employment areas above- vs. below-median density

Sector	OLS, mono-establishments				obs.
	\hat{A}	\hat{D}	\hat{S}	R^2	
	(1)	(2)	(3)	(4)	(5)
Food, beverages, tobacco	0.064 (0.004) *	0.968 (0.020)	0.006 (0.002) *	0.971	21,189
Apparel, leather	0.040 (0.011) *	1.353 (0.048) *	0.005 (0.005)	0.988	5,713
Publishing, printing, recorded media	0.170 (0.009) *	1.286 (0.047) *	-0.001 (0.004)	0.988	8,993
Pharmaceuticals, perfumes, soap	0.056 (0.068)	1.151 (0.161)	-0.003 (0.078)	0.744	1,016
Domestic appliances, furniture	0.117 (0.010) *	1.194 (0.045) *	0.004 (0.006)	0.990	6,172
Motor vehicles	0.086 (0.021) *	1.207 (0.159)	0.007 (0.017)	0.855	1,408
Ships, aircraft, railroad equipment	0.096 (0.043) *	1.165 (0.174)	0.000 (0.040)	0.852	964
Machinery	0.083 (0.005) *	1.052 (0.027)	-0.004 (0.003)	0.985	14,082
Electric and electronic equipment	0.083 (0.012) *	0.996 (0.055)	-0.002 (0.018)	0.948	5,550
Building materials, glass products	0.076 (0.013) *	1.125 (0.067)	0.003 (0.010)	0.973	3,048
Textiles	0.064 (0.013) *	1.079 (0.057)	0.005 (0.010)	0.920	3,275
Wood, paper	0.084 (0.011) *	1.135 (0.041) *	-0.002 (0.005)	0.989	5,627
Chemicals, rubber, plastics	0.073 (0.011) *	1.125 (0.039) *	0.005 (0.005)	0.969	5,119
Basic metals, metal products	0.069 (0.005) *	1.055 (0.023) *	0.001 (0.002)	0.993	13,911
Electric and electronic components	0.076 (0.023) *	0.993 (0.077)	-0.003 (0.031)	0.938	2,487
Consultancy, advertising, business services	0.190 (0.005) *	1.116 (0.021) *	-0.004 (0.002)	0.983	35,738
All sectors	0.091 (0.002) *	1.226 (0.009) *	0.001 (0.001)	0.997	134,275

*: for \hat{A} and \hat{S} significantly different from 0 at 5%, for \hat{D} significantly different from 1 at 5%.

Column (1) in table 1 reports our estimates of A . They are all positive. Statistical significance at the 5 percent level is marked with an asterisk next to the bootstrapped standard errors reported in parenthesis. All of our estimates for A except one are significant at 5 percent. When considering all sectors, we find $\hat{A} = 0.091$. This, on its own, implies an increase in mean productivity of $e^{0.091} - 1$, or 9.5 percent, in denser employment areas relative to less dense ones.

Column (2) in table 1 reports our estimates of D . In seven sectors our estimate of D is statistically significantly different from one. In all those cases, the estimated coefficient is above one. There is thus a tendency for the distribution of firm log productivity to be more dilated in denser employment areas for half the sectors and for all sectors combined. For all sectors, we find $\hat{D} = 1.226$. Dilating the log productivity distribution in employment areas with below-median density by this

value, shifting it by $\hat{A} = 0.091$ and left truncating a share $\hat{S} = 0.001$ of its values, results in a predicted productivity advantage of 9.7 for firms at the mean, 4.8 percent for firms at the bottom quartile, and 14.4 percent for firms at the top quartile. These are very close to the differences in the empirical distributions for employment areas with above- and below-median density, which are 9.7, 4.8, and 13.9 percent for the mean, bottom quartile, and top quartile respectively.¹⁶

Taken together, these estimates of A and D suggest that agglomeration economies, or more generally productive advantages shared by firms, are stronger in denser employment areas than in less dense areas. In our model, the extent to which there are common productivity advantages from larger cities is closely related to the extent to which interactions are local or global (national in this case). Our results are consistent with a situation where interactions are quite local. This matches the existing empirical literature (see Rosenthal and Strange, 2004).

Column (3) in table I reports our estimates of S . There is only one sector (food beverages and tobacco) with a positive and significant value for S , although this value is small at 0.006. In all other sectors, the point estimate for S is not significantly different from zero. This lack of significance is not due to imprecise estimates. On the contrary, in all sectors except pharmaceuticals, perfumes, and soap the standard errors for S are small, like the standard errors for A . These results provide strong evidence that there are no differences between denser and less dense employment areas in the truncation of the distribution of firm log productivity. Market selection appears to have a similar intensity across employment areas in France irrespective of their employment density.

Finally, column (4) in table I reports a pseudo- R^2 as defined in section 3. It measures how much of the mean squared quantile difference between the distributions of firm log productivity in denser and less dense employment areas is explained by our three parameters. The fit is very good. For all sectors together, virtually all the differences between the distributions of log productivity between denser and less dense employment areas is explained. For 13 out of 16 individual sectors, the pseudo- R^2 is above 0.900.

To summarise, firms are more productive in denser employment areas. However, this is not because tougher competition makes it more difficult for the least productive firms to survive. The productivity advantages of large cities arise because all firms see their productivity boosted, and in about half of the sectors this increase in productivity is strongest for the most productive firms.

Constrained specifications

We now explore to what extent it is important to estimate all three parameters A , D , and S by comparing our baseline results with constrained specifications. First, we study the importance of allowing more productive firms to benefit more from denser cities by estimating a simpler specification where all firms benefit equally and comparing it with our baseline. The first three columns of table II report estimates of A , S , and a pseudo- R^2 when we impose the restriction

¹⁶The difference between the 0.095 increase in mean TFP that we obtain from using $\hat{A} = 0.091$ alone and this 0.097 comes from applying the point estimate of the truncation parameter $\hat{S} = 0.001$, which raises mean TFP relative to employment areas with below-median density by 0.001, value that gets diluted by $D = 1.226$. For the bottom and top deciles of the distribution, we find estimated productivity advantages of 0.3 percent and 20.2 percent, respectively.

Table II: Constrained specifications, employment areas above- vs. below-median density

Sector	OLS, mono-establishments							obs.
	\hat{A}	\hat{S}	R^2	\hat{A}	R^2	\hat{S}	R^2	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Food, beverages, tobacco	0.062 (0.005)*	0.008 (0.002)*	0.957	0.071 (0.005)*	0.846	0.035 (0.009)*	0.468	21,189
Apparel, leather	0.093 (0.027)*	-0.045 (0.042)	0.462	0.046 (0.010)*	0.144	-0.012 (0.003)*	0.108	5,713
Publishing, printing, recorded media	0.202 (0.012)*	-0.031 (0.017)	0.844	0.167 (0.009)*	0.725	-0.000 (0.011)	0.001	8,993
Pharmaceuticals, perfumes, soap	0.072 (0.047)	-0.009 (0.036)	0.522	0.047 (0.047)	0.124	-0.007 (0.016)	0.252	1,016
Domestic appliances, furniture	0.131 (0.016)*	-0.006 (0.015)	0.832	0.123 (0.011)*	0.811	0.019 (0.021)	0.029	6,172
Motor vehicles	0.103 (0.019)*	-0.006 (0.010)	0.682	0.095 (0.021)*	0.653	0.005 (0.011)	0.015	1,408
Ships, aircraft, railroad equipment	0.111 (0.030)*	-0.009 (0.015)	0.690	0.096 (0.032)*	0.594	-0.002 (0.018)	0.004	964
Machinery	0.088 (0.006)*	-0.009 (0.004)*	0.963	0.079 (0.006)*	0.902	0.004 (0.011)	0.014	14,082
Electric and electronic equipment	0.083 (0.013)*	-0.002 (0.004)	0.948	0.081 (0.012)*	0.941	0.025 (0.014)	0.147	5,550
Building materials, glass products	0.085 (0.018)*	-0.004 (0.014)	0.822	0.081 (0.014)*	0.809	0.005 (0.010)	0.037	3,048
Textiles	0.071 (0.015)*	-0.001 (0.006)	0.821	0.070 (0.015)*	0.821	0.012 (0.011)	0.115	3,275
Wood, paper	0.095 (0.011)*	-0.013 (0.006)*	0.859	0.081 (0.012)*	0.726	-0.001 (0.007)	0.004	5,627
Chemicals, rubber, plastics	0.083 (0.011)*	-0.002 (0.005)	0.821	0.080 (0.012)*	0.810	0.006 (0.006)	0.043	5,119
Basic metals, metal products	0.072 (0.005)*	-0.002 (0.003)	0.957	0.070 (0.005)*	0.950	0.006 (0.005)	0.059	13,911
Electric and electronic components	0.076 (0.017)*	-0.003 (0.007)	0.937	0.072 (0.018)*	0.922	0.003 (0.013)	0.031	2,487
Consultancy, advertising, business services	0.208 (0.008)*	-0.018 (0.006)*	0.936	0.184 (0.006)*	0.883	0.052 (0.038)	0.044	35,738
All sectors	0.115 (0.004)*	-0.019 (0.004)*	0.721	0.093 (0.002)*	0.596	-0.001 (0.001)	0.003	134,275

*: significantly different from 0 at 5%.

$D = 1$ (no difference in the strength of dilation between denser and less dense employment areas). Our restriction $D = 1$ does not change the interpretations of A and S .

In column (1), our estimates of A are always positive. They are significantly different from zero in all cases but one. For all sectors we find a value $\hat{A} = 0.115$, which implies a 12.2 percent productivity increase. In column (2), for 12 sectors out of 16, S is not statistically different from zero. It is negative and significant in three sectors and for all sectors pooled together. It is positive and significant in one sector only. In all cases, however, S remains small. Our measure of fit in column (3) is also good.

While these results are consistent with the findings of table I, a more detailed comparison between tables II and I reveals that it is important to estimate D and allow for more productive firms to benefit more from denser cities. When one fails to do so by imposing $D = 1$, estimates

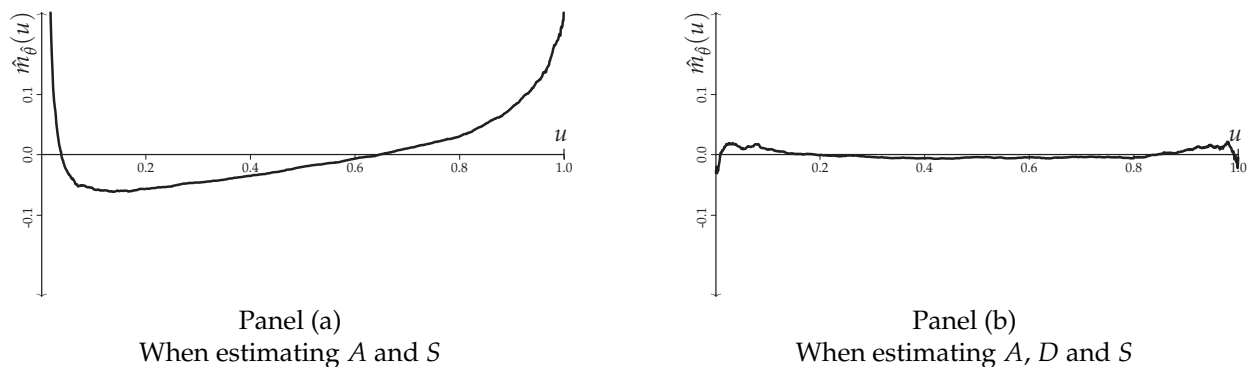


Figure 4: Estimation errors by quantile

of A and S become biased as they attempt to approximate a dilation. In particular, when we do not allow for $D > 1$, we tend to overestimate A and underestimate S (the latter even becoming negative in several cases). It is also clear from the comparison of tables II and I that the fit is better when considering A , D , and S instead of only A and S . Unsurprisingly, the improvement in the fit is strongest for those sectors with strong dilation. For instance, in apparel and leather, the pseudo- R^2 goes from 0.462 to 0.988 when adding D to the estimation.

Panels (a) and (b) of figure 4 provide further insight into this specification issue. The graph in panel (a) plots, for all sectors combined, the values of $\hat{m}_{\hat{\theta}}(u)$ coming from the bottom row of table II. That is, the figure plots for each quantile (given by a point on the horizontal axis) the difference between its value in the distribution of log productivity for denser areas and the value that results from shifting and truncating the distribution of log productivity in less dense areas using the estimated values of A and S when D is constrained to unity. Two features of panel (a) are noteworthy. First, errors for the first few quantiles are positive before quickly becoming negative above the first two percent of quantiles. This is due to the small negative value estimated for S ($\hat{S} = -0.019$), which leads to a bad fit at the very bottom of the distribution even if it helps improve the overall fit. Second, beyond those very first quantiles, there is a marked pattern where errors tend to be negative for the lower quantiles and positive for the higher quantiles. This indicates that, by forcing all establishments to have the same productivity boost from locating in a denser employment area, we are giving establishments at the lower end of the productivity distribution too large a boost (so the lower quantiles in the actual distribution for denser areas fall below those quantiles in the transformed distribution for less dense areas). At the same time, we are giving establishments at the upper end of the productivity distribution too small a boost. In other words, the figure indicates that more productive establishments benefit more from being in denser areas.

The graph in panel (b) plots, for all sectors combined, the values of $\hat{m}_{\hat{\theta}}(u)$ coming from the bottom row of table I. That is, the figure plots for each quantile the difference between its value in the distribution of log productivity for denser areas and the value that results from shifting, dilating, and truncating the distribution of log productivity in less dense areas using the estimated values of A , D , and S . Estimation errors are greatly reduced relative to those of panel (a). Allowing for dilation yields $\hat{S} = 0.001$ instead of $\hat{S} = -0.019$, which eliminates the large positive errors for the very first quantiles. It also eliminates the clear upward-sloping pattern apparent in panel (a).

In fact, errors in panel (b) are tiny everywhere except for a little wiggle at the both extremes, where productivity values are more scattered and the fit between the distributions loses precision.

We next impose additional restrictions to our specification by estimating either A alone or S alone. Columns (4) and (5) of table II report estimates of A and a pseudo- R^2 when we impose the restrictions $D = 1$ (no difference in dilation between denser and less dense employment areas) and $S = 0$ (no difference in truncation). Unsurprisingly given how close to zero the estimates of S in column (2) are, the estimates of A in column (4) are close to those in column (1). Column (5) reports the corresponding pseudo- R^2 and shows that for most sectors the fit does not deteriorate too much relative to column (3).

Columns (6) and (7) of table II report estimates of S and a pseudo- R^2 when we impose the restrictions $D = 1$ and $A = 0$ (no common productivity advantages for denser employment areas). In each and every case the estimate for S in column (6) is larger than or equal to its corresponding estimate in column (2). This suggests that if we do not allow denser areas to have common productivity advantages, we pick up part of their effects as variation in selection. Column (7) reports the pseudo- R^2 . A comparison with column (3) shows that the fit deteriorates substantially in all sectors but one. Overall, the results of columns (4)-(7) reinforce those of columns (1)-(3) by underscoring the robustness of our finding that there are no sizeable differences in left truncation between denser and less dense employment areas. This indicates that selection does not play a major role in explaining the productive advantages of denser areas. Instead, the entire log productivity distribution in denser areas is right-shifted and dilated relative to the distribution in less dense areas. This indicates that there are substantial productivity benefits for all firms in denser areas that are even stronger for more productive firms.

Robustness to alternative measures of TFP and samples of establishments

One might ask whether our results are robust to using alternative approaches to estimate TFP. While OLS is arguably the most transparent method to estimate TFP, it does not account for the possible simultaneous determination of productivity and factor usage. The top panel of table III reports results for all sectors combined using two approaches that account for this simultaneity, proposed by Olley and Pakes (1996) and Levinsohn and Petrin (2003), as well as a simple cost-share approach.

To ease comparisons, the first row of results reports the same OLS estimates as the last row of table I. The next row reports results for the same estimation of A , D , and S using the approach proposed by Olley and Pakes (1996) instead of OLS. The Olley-Pakes estimate of A , 0.087, is very close to its corresponding OLS value of 0.091. The estimates of S are also very similar. Finally, the estimate of the dilation parameter D is smaller when using Olley-Pakes: 1.087 against 1.226 with OLS. Estimating TFP using the method proposed by Levinsohn and Petrin (2003) in the third row of results in table III yields estimates that are very similar to those of Olley-Pakes TFP. The fourth row reproduces the estimation of A , D , and S when the underlying TFP is estimated using a simple

Table III: Robustness, alternative estimation methods

Method	\hat{A}	\hat{D}	\hat{S}	R^2	obs.
	(1)	(2)	(3)	(4)	(5)
all sectors, mono-establishments					
Ordinary Least Squares	0.091 (0.002) *	1.226 (0.009) *	0.001 (0.001)	0.997	134,275
Olley-Pakes	0.087 (0.006) *	1.087 (0.040) *	0.003 (0.003)	0.983	56,130
Levinsohn-Petrin	0.098 (0.003) *	1.112 (0.010) *	-0.000 (0.001)	0.996	99,145
Cost shares	0.084 (0.002) *	1.200 (0.010) *	0.002 (0.001)	0.983	134,275
all sectors, all establishments					
Ordinary Least Squares	0.095 (0.002) *	1.202 (0.011) *	0.000 (0.001)	0.998	162,765
Olley-Pakes	0.090 (0.007) *	1.152 (0.038) *	0.008 (0.003) *	0.995	73,974
Levinsohn-Petrin	0.114 (0.005) *	1.092 (0.016) *	-0.002 (0.002)	0.995	122,489
Cost shares	0.083 (0.003) *	1.151 (0.016) *	0.000 (0.001)	0.992	162,765

*: for \hat{A} and \hat{S} significantly different from 0 at 5%, for \hat{D} significantly different from 1 at 5%.

cost-share approach, which are also similar.¹⁷

Overall the differences between OLS and these alternative techniques to estimate TFP are small. The estimates of A and S are almost identical in all cases. The estimates of D show slightly more variation but remain above one and significant in all cases.¹⁸ While we do not report detailed sectoral results for these alternative TFP estimations, we note that they are close to the results reported in table I.

Since data for value added and capital is only available at the firm level, we have so far restricted the sample to firms with a single establishment. The bottom panel of table III replicates the same four estimations of A , D and S as the first panel but this time considering all establishments, including those that belong to firms with establishments in multiple locations, using the methodology explained in section 4. For OLS TFP, the results are the same as those with mono-establishment firms, except for slightly stronger agglomeration and slightly less dilation in denser employment areas. The next three rows report results for the alternative approaches to TFP estimations as described above. The point estimates are extremely close to but less precisely estimated than those obtained for mono-establishment firms and reported in the first panel of the same table. Another minor difference is that, when using Olley-Pakes, the tiny amount of truncation (the estimated parameter is $\hat{S} = 0.008$) is statistically significant. The point estimates are also close to those

¹⁷We do not use the method proposed by Syverson (2004) using instrumented cost shares. This approach, which uses local demand shocks as instruments, is valid only for industries with localised markets. It is not suitable for a broad cross-section of sectors nor when pulling all sectors together.

¹⁸This appears to be due to the estimation technique and not to the sample of establishments used. Estimating OLS TFP on the same sample used to estimate Olley-Pakes TFP (56,130 instead of 134,275 establishments) results in: $\hat{A} = 0.087$, $\hat{D} = 1.206$, and $\hat{S} = 0.004$.

Table iv: Robustness, alternative spatial units

Comparison	OLS, all sectors, mono-establishments				
	\hat{A}	\hat{D}	\hat{S}	R^2	obs.
	(1)	(2)	(3)	(4)	(5)
Employment areas, above vs. below median density	0.091 (0.002) *	1.226 (0.009) *	0.001 (0.001)	0.997	134,275
Employment areas, top vs. 3rd density quartile	0.116 (0.003) *	1.222 (0.012) *	0.001 (0.001)	0.996	76,793
Employment areas, 3rd vs. 2nd density quartile	0.022 (0.002) *	1.075 (0.012) *	-0.000 (0.001)	0.988	68,858
Employment areas, 2nd vs. bottom density quartile	0.026 (0.003) *	1.025 (0.012) *	0.000 (0.001)	0.983	57,481
Employment areas, above vs. below median density, conditional on high market potential	0.123 (0.003) *	1.296 (0.015) *	0.001 (0.001)	0.996	74,242
Cities, pop. > 200,000 vs. pop. < 200,000	0.087 (0.002) *	1.241 (0.009) *	0.000 (0.001)	0.998	134,275
Paris vs. cities with pop. 1–2 million	0.131 (0.004) *	1.187 (0.020) *	-0.001 (0.001)	0.995	46,935
Cities with pop. 1–2 million vs. pop. 200,000–1 million	0.038 (0.005) *	1.042 (0.021) *	0.003 (0.002)	0.964	36,582
Cities with pop. 200,000–1 million vs. pop. < 200,000	0.000 (0.003)	1.077 (0.011) *	-0.002 (0.001) *	0.953	87,341
Paris vs. Lyon (pop. 10,381,376 vs. 1,529,824)	0.096 (0.006) *	1.222 (0.030) *	0.000 (0.002)	0.987	41,336
Lyon vs. Nantes (pop. 1,529,824 vs. 621,228)	0.047 (0.010) *	0.989 (0.049)	-0.001 (0.005)	0.889	6,818
Nantes vs. Bayonne (pop. 621,228 vs. 65,944)	0.054 (0.020) *	1.059 (0.097)	0.002 (0.018)	0.885	1,905

*: for \hat{A} and \hat{S} significantly different from 0 at 5%, for \hat{D} significantly different from 1 at 5%.

obtained with OLS TFP. Overall we conclude that neither the sample of establishments we use nor the specific method we implement to estimate TFP have much bearing on our results.

Robustness to alternative spatial units and measures of local scale

One might question whether our results are driven by our use of employment areas as spatial units and being above or below median employment density as criterion for grouping them. Employment areas are natural units within which to explore agglomeration effects because they closely match local labour markets. Employment areas are less likely to provide good approximations for markets for final goods and thus might be less appropriate when searching for market selection effects.¹⁹ As for our grouping criterion, density has often been used by past research (e.g., Ciccone and Hall, 1996, Combes *et al.*, 2010) but it is by no means the only measure of local scale. A comparison of places above- and below-median density is also natural but might hide

¹⁹Recall nonetheless that our objective is to understand whether local differences in productivity are driven by agglomeration or market selection. A complete search for whether market selection effects can be observed at any spatial scale is of course beyond the scope of this paper.

more subtle differences. To check whether our main results are robust to our choice of spatial units and criterion for grouping them, we can replicate them using alternative units like urban areas, alternative measures of local scale such as population size, and finer groupings such as groupings by quartile or comparisons of particular places.

Table IV reports a number of results for alternative spatial units and grouping criteria. The first row of results reproduces again our main results comparing employment areas with above- and below-median employment density. The next three rows of table IV divide French employment areas into four groups (by density quartiles) instead of just two. While the results generally confirm our main results, they highlight a large gap for A and D between the fourth density quartile, which contains the densest employment areas and the third density quartile. For A and D , the differences between the third and second density quartile or between the second and first density quartile are much smaller but remains nonetheless statistically significant. These differences in estimates of A and D across quartiles reflect the distribution of density across employment areas in France. The average density of the employment areas in the second quartile is slightly more than twice that of employment areas in the first quartile. The average density in the third quartile is slightly less than twice that in the second quartile. By contrast the average density in the top quartile is nearly 12 times than in the third quartile. It is interesting to note that the estimates of A in the different quartiles are roughly proportional to those ratios. This is consistent with panel (a) of Figure 1 which is suggestive of a log linear relationship between density and mean TFP . These finer results also confirm the absence of selection in all cases.

One may worry that the local density of employment may be strongly correlated with better access to product markets. To verify that our results continue to hold even after factoring out the higher market potential of denser areas, we construct for each area a simple market potential index by summing the density of its neighbours weighted by their inverse distance. The fifth row of results in table IV repeats the same estimation as the first row but considers only employment areas with above-median market potential. This yields similar results.

The sixth row of results in table IV repeats again the estimation of the first row, but compares urban areas with over 200,000 people and urban areas with less than 200,000 people and rural areas instead of employment areas with above and below median employment density. Urban area boundaries are drawn to capture cities whereas employment area boundaries are drawn to capture local labour markets on the basis of commuting patterns. While the total number of areas is roughly similar (341 contiguous employment areas instead of 364 urban areas and the rural areas that surround them), differences are substantial. For instance, Greater Paris is classified as a single urban area but is made up of 16 separate employment areas. Nevertheless, the estimated coefficients for A , D , and S are almost identical. The fit is also excellent. In the separate web appendix, we report detailed sector by sector results for French urban areas to compare with those of table I. Results are again similar. Splitting urban areas into four categories in the seventh to ninth rows of table IV, as we do with employment areas in the second to fourth rows, also gives similar results. We conclude that grouping cities according to population size or employment areas according to employment density yields very similar results.

Grouping areas, as we have done so far, is useful because it ensures that we have enough

Table v: Noisy truncation, simulation results

Standard dev. of noise relative to standard dev. log of TFP distribution	Simulated log-normal TFP distribution with added noise				
	\hat{A}	\hat{D}	\hat{S}	R^2	obs.
	(1)	(2)	(3)	(4)	(5)
0%	0.100 (0.021) *	1.200 (0.015) *	0.100 (0.009) *	1.000	93,102
5%	0.100 (0.015) *	1.199 (0.010) *	0.100 (0.005) *	1.000	93,102
10%	0.102 (0.015) *	1.197 (0.010) *	0.099 (0.005) *	1.000	93,102
20%	0.115 (0.015) *	1.185 (0.010) *	0.092 (0.005) *	0.999	93,102
30%	0.137 (0.014) *	1.166 (0.010) *	0.081 (0.005) *	0.999	93,102

The simulations use parameter values $A = 0.100$, $D = 1.200$, and $S = 0.100$.

*: for \hat{A} and \hat{S} significantly different from 0 at 5%, for \hat{D} significantly different from 1 at 5%.

observations to estimate parameters accurately and reduces the impact of idiosyncrasies associated with any particular areas. Nevertheless, it may be instructive to look at a few examples. The last three rows of table iv perform pairwise comparisons of individual cities that are illustrative of our general results. The four cities used in these comparisons are Paris (the largest, with a population above 10 million), Lyon (the second largest, with a population around 1.5 million), Nantes (about half a million), and Bayonne (a smaller city, with a population below 100,000). Although the number of observations becomes small for the comparison between Nantes and Bayonne, the estimate of A remains significant. A trebling of population between Nantes and Lyon is associated with a 5 percent increase in average TFP. The productivity gap reflected in the estimate of A for the comparison between Paris and Lyon is of the same magnitude, once we account for the fact that Paris is larger than Lyon by a factor of nearly seven. As also expected in light of previous results, there are no differences in the strength of selection. Note also that the fit deteriorates as the number of observations becomes small for comparisons involving smaller cities.

6. Discussion

Detecting truncation with noise

To assess how much the distribution of log productivity in denser areas is shifted, dilated, and truncated relative to the same distribution in less dense areas, we must use an estimate of the productivity of each establishment rather than its true value. If productivity is estimated with noise, truncation may not be immediately apparent from the distribution of measured log TFP as market selection eliminates establishments below some threshold of true productivity. In this subsection, we report simulation results showing that our methodology is able to identify truncation accurately when it is present in the distribution of true productivity, even if TFP is estimated with a substantial amount of noise.

To evaluate the effects of noise in measured TFP on our results, we consider a hypothetical population of establishments. For establishments located in less dense areas, the unit labour requirement h is assumed to be drawn from a log-normal distribution with mean zero and unit variance, implying that true productivity is also log-normally distributed with mean zero and unit variance. We use a log-normal distribution for simulations in this section because, as shown in appendix A, it provides a good approximation to the empirical TFP distribution. In results not reported here, we have experimented with other distributions and obtained very similar results. For establishments located in denser areas, the distribution of true log productivity is shifted, dilated and truncated relative to that in less dense areas. We assume that the shift and dilation parameters are $A = 0.100$ and $D = 1.200$, to match (rounded to the first decimal) our preferred empirical estimates. For the selection parameter, we assume $S = 0.100$. This value is much higher than our preferred estimate $\hat{S} = 0.001$ because we are interested in checking whether actual left-truncation could be missed by our approach due to noisy TFP estimation. We introduce noise in the productivity estimation by making observed log TFP be the sum of true log productivity and a random error drawn from a normal distribution with zero mean and variance ς^2 .

Table v reports estimates of A , D , and S and their standard errors, using 1000 simulated samples.²⁰ Each row in the table corresponds to a different magnitude of the noise introduced in TFP, measured in terms of how large is the standard deviation of the noise relative to the standard deviation of the entire distribution of log productivity (equal to ς given unit variance for the distribution of true productivity). The first row of results confirms that when true productivity is observed ($\varsigma = 0$), we recover the true parameters used for the simulations: $\hat{A} = 0.100$, $\hat{D} = 1.200$, and $\hat{S} = 0.100$. The next two rows show that for small to moderate noise in measured TFP ($\varsigma = 0.05$ or 0.10 , equivalent to having a standard deviation of the noise equal to 5 or 10% of the standard deviation of the distribution of true log productivity), we recover almost exactly the true parameters used for the simulations. In the last two rows, for $\varsigma = 0.20$ or 0.30 , we can see an upward bias in \hat{A} and a downward bias in both \hat{D} and \hat{S} . However, these values of ς correspond to a very high level of noise in TFP estimates. When $\varsigma = 0.30$, the standard deviation of the noise is 30% of the standard deviation of the distribution of log productivity, and the 95% confidence interval for an establishment with observed mean TFP is between the 17st and the 83th percentile. Even then, the estimate $\hat{S} = 0.081$ associated with $\varsigma = 0.30$ remains nearly two orders of magnitude higher than our preferred estimate of S when using actual data, and is significantly different from zero. This shows that even if TFP is estimated with a substantial amount of noise, our methodology is still able to detect truncation when it is present in the distribution of true productivity.

Product-level selection

In our model, firms produce a single differentiated product, while, in reality, many firms produce multiple products. This raises the question of whether with selection at the level of individual products we would still observe left-truncation of the log TFP distribution for firms. To show that

²⁰Each sample begins with 100,000 simulated observations equally split between denser and less dense areas. The 93,102 observations reported in the table reflect the elimination through selection of 10% of observations in denser areas, and the trimming of 1% of observations at both extremes, as in our baseline results, to remove outliers.

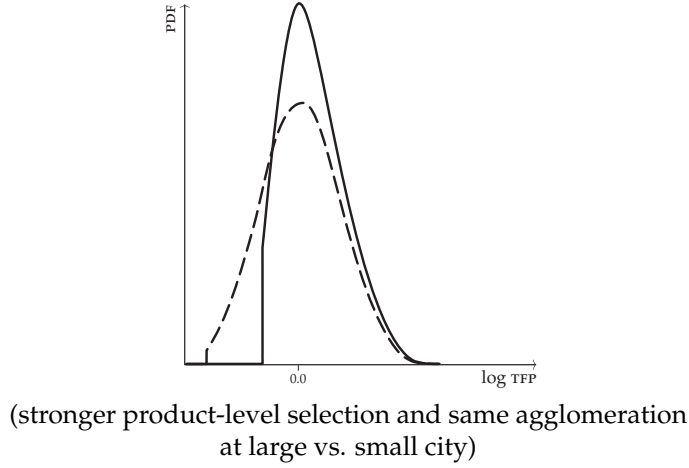


Figure 5: Log TFP distribution in large (solid) and small cities (dashed) with product-level selection

this is indeed the case, we now extend our model to allow for multi-product firms. In doing so, we combine elements of two recent models of selection with multi-product firms. These are Bernard, Redding, and Schott (2010) and Mayer, Melitz, and Ottaviano (2011), although in the case of the former we remain closer to the static version in Bernard, Redding, and Schott (2006).

Following Mayer *et al.* (2011), to ensure that the assumption of monopolistic competition can be maintained, let us assume that individual firms produce a countable number of products up to a maximum of K . In equilibrium firms will nevertheless differ in terms of how many products they make and the productivity level they can achieve for each of them. Following Bernard *et al.* (2006), let us assume that the unit labour requirement for a product is now the product of two components. The first component, h^0 , is common to all products sold by the firm (Bernard *et al.*, 2006, call this ‘ability’) and drawn from a distribution with known differentiable probability density function $g(h^0)$. The second component, h^k , is specific to product k (Bernard *et al.*, 2006, call this ‘expertise’) and drawn from a distribution with known differentiable probability density function $r(h^k)$. Since we simply wish to show that differences in selection still get reflected in differences in left-truncation, we focus on a case of two cities of different sizes in which product-market competition is local and interactions are global. Without differences in agglomeration across locations, we can set $A_i = 0$ and $D_i = 1, \forall i$. In that case, TFP at the product level is $1/(h^0 h^k)$. As in our baseline model, market selection still implies that firms cannot find positive demand for products for which their unit labour requirement is above \bar{h}_i . To compute log TFP at the firm level, we need to take into account that firms produce only a subset of their potential range of products and that each of them is produced in different quantities. From equations (4) and (5), product-level output is $N_i(\bar{h}_i - h^0 h^k)/2\gamma$. Log TFP at the firm level is then the weighted average of TFP at the product level:

$$\phi_i(h^0, h^1, \dots, h^K) = \ln \left(\frac{\sum_{h^k | h^k \leq \bar{h}_i} \frac{\bar{h}_i - h^0 h^k}{h^0 h^k}}{\sum_{h^k | h^k \leq \bar{h}_i} \bar{h}_i - h^0 h^k} \right) .$$

Figure 5 plots the distributions of log TFP, under the assumption that there are differences only in the strength of selection across cities, keeping the rest of the model as in the baseline case. Thus,

it corresponds to a version of panel (a) in figure 3 with multi-product firms and product-level selection.²¹ The key feature to note is that differences in the strength of selection across cities still get reflected in differences in truncation. An individual firm's TFP has a lower bound at the TFP of its weakest product, which in turn has a lower bound at the cut-off resulting from local product-market competition. Thus, if multi-product firms arise when firms expand beyond the product where their expertise is highest, stronger product-level selection should still result in left-truncation of the firm-level TFP distribution.²²

Two other features of this extension are also worth noting. First, firms in the large city end up selling fewer products for any given set of draws h^0, h^1, \dots, h^K . This is because, in the face of tougher competition, firms do not expand their product range as far beyond products where their 'expertise' is highest. Hence, firms of any given TFP level produce fewer products on average in the large city. Thus, showing that the number of products, conditional on TFP, does not decrease with the size or density of areas would be an additional piece of evidence against differences in the strength of selection. Unfortunately, the data required to do this is not available for France.

Second, if we consider both differences in selection and in agglomeration in this multi-product extension, differences in truncation still reflect differences in selection. However, if contrary to our empirical results one were to find differences in truncation in the data ($\hat{S} \neq 0$), interpreting \hat{A} and \hat{D} would become more difficult. In this case, a firm with a given set of draws h^0, h^1, \dots, h^K ends up producing fewer products if based in the large city. Thus, its measured TFP is higher. Since we find S to be not significantly different from zero, this is not a worry for us.

Nevertheless, a rough way to control for the number of products is to use firm size. Table vi repeats our baseline estimation for subsets of establishments of different size. We can see that S is very close to zero across all rows regardless of the firm size class. Furthermore, the estimate of A remains positive and increases gradually with size. This naturally follows from the positive association between firm size and productivity and our finding that more productive firms benefit more from agglomeration (i.e., our finding of significant dilation in the overall distribution).

The consequences of unobserved prices

As is often the case in the estimation of production functions, we do not observe prices in the data. Thus, we must estimate productivity by studying how much value (instead of physical output) an

²¹The figure is drawn under the assumptions that $K = 5$ and $S = 0.2$, and that both $g(h^0)$ and $r(h^k)$ are normal distributions with mean 0 and variances 1 and 0.2 respectively. This already makes the multi-product component much more prominent than it appears to be in reality, as almost all firms become multi-product to different degrees, and multi-product firms produce 4.8 products on average. According to Bernard *et al.* (2010), in the United States 39% of firms are multi-product and they produce 3.5 products on average.

²²That is, product-level selection implies firm-level selection. However the converse is not true in general. For instance, if there are complementarities, firms could maintain weak products because of positive effects on the rest of their product range, a form of economies of scope. In this case, we could find firm-level selection even without product-level selection.

Table VI: Estimation results by firm size, employment areas above- vs. below-median density

Establishments	Employment	OLS				obs.
		A	D	S	R^2	
		(1)	(2)	(3)	(4)	(5)
1	1–5	0.034 (0.003) *	1.176 (0.005) *	0.002 (0.001) *	0.992	198,167
1	6–10	0.075 (0.003) *	1.173 (0.012) *	0.001 (0.001)	0.998	69,901
1	11–20	0.113 (0.004) *	1.223 (0.018) *	0.001 (0.001)	0.993	30,694
1	21–100	0.134 (0.004) *	1.305 (0.021) *	-0.000 (0.002)	0.997	29,107
1	> 100	0.180 (0.013) *	1.401 (0.091) *	0.001 (0.007)	0.992	4,578
> 1	Any	0.117 (0.008) *	1.125 (0.033) *	-0.000 (0.004)	0.996	28,491

*: for \hat{A} and \hat{S} significantly different from 0 at 5%, for \hat{D} significantly different from 1 at 5%.

establishment can produce with given inputs.²³

Although using value added to estimate productivity may affect TFP estimates, it is important to note that it will not bias our estimate of the selection parameter S , even if markups are systematically related to city size. The value of log TFP at which each distribution might be left-truncated can be different, since price markups are included in log TFP estimates. However, recall that S is the share of establishments in the small city distribution that are truncated out of the large city distribution, and is thus not affected by this. Whether A and D are biased or not depends on the value of S , as we show in appendix D. Finding $\hat{S} = 0$ implies that price markups are not systematically related to city size, and in this case A and D are also unbiased.

Given this finding, we now discuss more direct evidence that producer prices are indeed not systematically related to city size. Our model suggests that, if product markets are not sufficiently integrated, producer prices will be lower in large cities. If markets are closely integrated instead (or if firms use uniform delivered pricing within France), prices will not be systematically related to city size, implying no differences in the strength of selection. Unfortunately, we do not have direct evidence on the actual relationship between producer prices and city sizes. Some papers look at the variation in consumer prices across cities of different sizes and suggest that prices may increase with city size (e.g., Albouy, 2008). However, two problems prevent us from using this evidence against firm selection models. First, differences in consumer prices may reflect differences

²³Even if prices were observed, it would be unclear whether higher prices reflect higher markups or higher quality. The literature suggests two solutions that work only for specific industries. One can focus on homogeneous goods for which quantities are directly observed, like ready-made concrete (Syverson, 2004, Foster, Haltiwanger, and Syverson, 2008). Alternatively, one can focus on industries with localised markets for which direct measures of quality are available, like newspapers and restaurants (Berry and Waldfogel, 2010). For industries that do not meet these characteristics, a third alternative is to consider detailed product-level information, including prices, to recover the price markup of firms and back up their output-based productivity (De Loecker, 2011). However, to disentangle whether higher prices reflect larger markups or superior quality, one still has to make specific assumptions about how quality is produced and about the functional form of demand.

in retail costs rather than differences in markups. While data on producer prices is unavailable, we estimate that differences in retail costs account for over 40% of differences in consumer prices within product categories in France.²⁴ Second, higher consumer prices in large cities may reflect the well-established fact that the wealthier households that are disproportionately located in large cities consume higher quality (and substantially more expensive) varieties, even within narrowly defined product categories and in the same store (Bils and Klenow, 2001, Broda, Leibtag, and Weinstein, 2009). Handbury and Weinstein (2010) is the first paper to study the relationship between prices and city sizes for a broad range of truly identical products, thanks to the use of a rich dataset based on Universal Product Code (UPC) scans. Even before controlling for differences in retail costs due to the land used and the amenities provided by stores in different areas, they find that there is no statistically significant relationship between prices within each UPC and city size. This evidence for the United States is consistent with our finding for France of \hat{S} close to 0.

A comparison with approaches based on summary statistics

A key innovation of our approach is that we use information from the entire distribution of firm log productivity to study to what extent average productivity is higher in denser cities because there is greater selection that eliminates the least productive firms or because there are productivity benefits that are, to some extent, shared by all firms. While this differs considerably from extant approaches in the literature, we can nonetheless relate our results to previous contributions studying just one of these two broad reasons on the basis of summary statistics.

Starting with Sveikauskas (1975), the empirical literature on agglomeration typically estimates the elasticity of some measure of average productivity, like average TFP, with respect to some measure of local scale, such as employment density or total population. More recent studies have paid particular attention to addressing two potential problems with that approach. First, more productive workers may sort into denser areas because of stronger preferences for the amenities typically found in those areas or because they benefit more from the productive advantages of higher density. The standard way to deal with this issue is to use detailed data on worker characteristics or even to exploit a panel to incorporate individual fixed effects in a regression of individual wages on city density (Glaeser and Maré, 2001, Combes, Duranton, and Gobillon, 2008). This issue of endogenous labour quality turns out to be important in practice. In light of this, we take advantage of having information on the hours worked by each employee in each establishment and their detailed occupational code, to incorporate detailed labour quality into our TFP estimation. A second identification issue is that productivity and density are simultaneously

²⁴Using confidential price data that underlie the French consumer price index (nearly 35,000 observations for April 2002), we calculate an elasticity with respect to city size of consumption prices within each of 373 product categories of 0.011. To estimate what percentage of this can be attributed to differences in retail costs alone, we use information reported in Betancourt and Gautschi (1996), which suggests that retail accounts on average for 35% of consumer prices in France. While we lack data for the importance of land in retail for France, Jorgenson, Ho, and Stiroh (2005) report a land share of 1.9% for retail in the United States and Chaney, Sraer, and Thesmar (2007) find that in other sectors this share is similar between France and the United States. Combes, Duranton, and Gobillon (2011) estimate the elasticity of unit land prices with respect to population in French urban areas to be around 0.8. A retail share of 35%, combined with a land share in retail of 1.9%, and an elasticity of land prices with respect to city population of 0.8 would account for about 48% ($0.35 \times 0.019 \times 0.8/0.011$) of the elasticity of consumer prices with respect to population.

determined. With localised natural advantage, some areas are more productive and, as a result, become denser. Starting with Ciccone and Hall (1996), the standard way to tackle this potential problem is to use instrumental variables when regressing average productivity on local size or density. The main finding is that reverse causality or simultaneity is only a minor issue in practice, including in France (Combes *et al.*, 2010).

We can relate our estimate of A , the common shift in log productivity of establishments in denser areas relative to their counterparts in less dense areas, to the findings of the literature that attempts to separate agglomeration economies from localised comparative advantage. We do this by first turning our estimate of A into an elasticity of average TFP with respect to density. An average employee in a French employment area with above-median density benefits from a density that is 2.8 log points higher than an average employee in an area with below-median density. This difference implies that our estimate of $\hat{A} = 0.091$ for all sectors combined in table 1 is equivalent to an (arc) elasticity of TFP with respect to employment density of $0.091/2.8 = 0.032$. When regressing mean TFP for French employment areas on employment density in those areas, Combes *et al.* (2010) find an elasticity of 0.035.²⁵ This elasticity captures the combined effect of agglomeration economies and localised natural advantage. When instrumenting density by long historical lags of population or by soil characteristics to isolate agglomeration effects, they estimate only a slightly lower elasticity of 0.029. While we recognise the strong identification assumptions behind this result (or others similar to it), it is nonetheless suggestive that agglomeration effects are behind most of the shift in the log productivity distribution between less dense and denser areas that we observe in the data.

Turning to market selection, existing approaches are harder to compare to ours. Like Syverson (2004), Melitz and Ottaviano (2008), and other models relating selection to market size, ours also predicts that tougher competition leads to a left truncation of the distribution of productivity in denser employment areas relative to less dense areas. Unfortunately, detecting left truncation on the basis of summary statistics such as the mean or variance of firm productivity is not straightforward. Greater left truncation increases average productivity, but so does agglomeration. Both selection and agglomeration can also explain an increase in the median or the bottom decile of local productivity. In the model of Syverson (2004), left truncation also implies a decrease in the variance of productivity. We note that this result depends crucially on distributional assumptions.²⁶ Furthermore, it is possible that the strength of both selection and agglomeration increases

²⁵It is worth noting that Combes *et al.* (2010) estimate the elasticity of TFP with respect to urban density using a measure of local TFP that weights firm-level TFP by firms' employment shares. This could lead to different results than using an unweighted average of TFP if denser cities had greater allocative efficiency, i.e., if relatively more resources were allocated to more productive firms in denser cities. However, this turns out not to be the case. A way to assess this more formally is, following Olley and Pakes (1996), to decompose employment-share weighted average TFP for each city into the sum of two components: unweighted average TFP and a cross-term measuring allocative efficiency. The correlation between the cross-term and urban density for French employment areas is 0.02 and not statistically significantly different from zero. This decomposition is interesting because it confirms that greater allocative efficiency is not behind the aggregate productivity advantage of denser cities.

²⁶The result that the variance of productivity decreases with left truncation holds in Syverson's model and, more generally, for productivity distributions with log-concave density. However, this result would be reversed if one considered instead a productivity distribution with log-convex density, such as the Pareto distribution commonly used in this literature (on the relationship between the variance of a left truncated distribution and log-concavity and log-convexity, see Heckman and Honore, 1990).

with employment density in certain sectors. Even if the shape of the distribution was such that truncation reduced dispersion, agglomeration could simultaneously increase dispersion through a dilation of the distribution, and thus make the separation of selection and agglomeration based on dispersion measures alone difficult. A key difference with our approach is that we consider simultaneously selection and agglomeration and look at all quantiles of the productivity distributions, so that we do not rely on particular distributional restrictions. Finally, Syverson focuses on one sector, ready-made concrete, chosen because of particular characteristics. We look instead at a broad cross-section of sectors.²⁷

Given these differences with existing approaches, a detailed comparison of results would not be informative. Instead, we can ask how large selection effects would need to be in our data to generate the differences in average productivity that we observe in the absence of any agglomeration economies. To conduct this exercise we solve for S , with $A = 0$ and $D = 1$, so as to match existing difference in mean productivity between denser and less dense employment areas. We find that to explain a difference in mean log TFP of 0.09 between areas with employment density above and below the median, S should be equal to 0.15. When doing the same calculation sector by sector we find that selection effects of similar magnitude would be needed to explain observed differences in mean productivity. Put differently, for selection effects to be the main force at play behind existing differences in average productivity across cities, they would need to be two full orders of magnitude larger than our current estimates.

7. Concluding comments

To assess the importance of firm selection relative other motives commonly considered by urban economists for explaining the productive advantages of larger cities, we nest a standard model of agglomeration with a generalised version of the firm selection model of Melitz and Ottaviano (2008). The main prediction of our model is that stronger selection in larger cities left-truncates the firm productivity distribution while stronger agglomeration right-shifts and dilates it. A similar prediction would emerge from a much broader class of models nesting selection and agglomeration plus localised natural advantage, provided the underlying distribution of firm productivity is the same everywhere and selection effects can be separated from agglomeration effects. An important benefit of our structural approach is that it allows for a tight parametrisation of the strength of both types of forces.

To implement this model on exhaustive French establishment-level data, we develop a new quantile approach that allows us to estimate a relative change in left truncation, shift, and dilation between two distributions. This approach is general enough that it could be applied to a broad set of issues involving a comparison of distributions. When implemented with distributions of firm log productivity, this quantile approach is fully consistent with our theoretical framework.

Our main finding is that selection explains none of the productivity differences across areas in France. The distribution of firm log productivity in denser French employment areas is remarkably

²⁷The approach developed in Del Gatto *et al.* (2008) also differs significantly from ours. They make distributional assumptions about productivity and assess whether more open sectors exhibit a smaller dispersion of productivity.

well described by taking the distribution of firm productivity in less dense French employment areas, dilating it, and shifting it to the right. This corresponds to there being some productivity advantages for all firms from locating in denser areas, which are particularly strong for those firms that are per se more productive. This result holds for the productivity distributions of firms across all sectors as well as most two-digit sectors when considered individually. This finding is also robust to the choice of zoning. Our bottom line is that the distribution of firm log productivity in areas with above-median density is shifted to the right by 0.091 and dilated by a factor of 1.226 relative to areas below median density. Firms in denser areas are thus on average about 9.7 percent more productive than in less dense areas. Because of dilation, this productivity advantage is only of 4.8 percent for firms at the bottom quartile and 14.4 percent for firms at the top quartile. On the other hand we find no difference between denser and less dense areas in terms of left truncation of the log productivity distribution.

These findings are interesting and raise a number of questions regarding future research. Most models of agglomeration economies can easily replicate a shift but far fewer imply a dilation (Durananton and Puga, 2004). In our model, dilation arises from a simple technological complementarity between the productivity of firms and that of workers. Such complementarity could arguably be generated from more subtle interactions between firms and workers (assuming for instance some heterogeneity among workers as well). Furthermore this type of complementarity might also have some interesting implications with respect to location choices for both firms and workers as well as implications regarding the dynamics of firm productivity and workers' career paths.

That there are no differences in market selection might seem surprising to some. The emphasis however should be on the word difference. The fact that distributions of firm log productivity all exhibit a positive skew would be consistent with some selection if the underlying distribution of productivity were symmetric (or negatively skewed). However such selection appears to take place everywhere in France with the same intensity. As shown by our model, this is consistent with the French market being highly integrated either because the cost of delivering goods to different locations does not differ much or because firms may choose to offer the same price in all areas. Different findings could certainly emerge when comparing different countries.

Furthermore, our finding of no difference in selection across places is consistent with the usual finding in the trade literature that trade liberalisation raises productivity mostly through selection. Poorly integrated markets might show big differences in the intensity of market selection whereas highly integrated markets might have very little. Any transition between these two states involves changes in selection. For instance, when a country liberalises its imports, many low productivity firms may be eliminated by stronger competition from foreign competitors. However, as trade liberalisation proceeds further, the toughness of competition and thus the strength of market selection will converge between the home and foreign countries. This end result of no large spatial differences in the strength of selection is what we find when comparing cities across France.

At a different spatial scale, we also suspect that for many consumer services selection could be stronger at a fine level of aggregation such as the neighbourhood. A new hairdresser on a stretch of street is likely to affect other hairdressers along that stretch through increased competition more than a new car producer will affect other car producers in the same city. In the latter case, producers

sell to consumers across the country, or even across the continent, and the main effects of co-location are thus the usual benefits of agglomeration economies (from sharing suppliers, having a common labour pool, or learning spillovers) rather than spatial differences in selection, which appear to be very small across a highly integrated market.

References

- Albouy, David. 2008. Are big cities really bad places to live? Improving quality-of-life estimates across cities. Working Paper 14472, National Bureau of Economic Research.
- Baldwin, Richard E. and Toshihiro Okubo. 2006. Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting. *Journal of Economic Geography* 6(3):323–346.
- Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2010. Productive cities: Sorting, selection, and agglomeration. Processed, University of Toronto.
- Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. Plants and productivity in international trade. *American Economic Review* 93(4):1268–1290.
- Bernard, Andrew B. and J. Bradford Jensen. 1999. Exceptional exporter performance: Cause, effect, or both? *Journal of International Economics* 47(1):1–25.
- Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2006. Multi-product firms and trade liberalization. Working Paper 12782, National Bureau of Economic Research.
- Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2010. Multiple-product firms and product switching. *American Economic Review* 100(1):70–97.
- Berry, Steven and Joel Waldfogel. 2010. Quality and market size. *Journal of Industrial Economics* 58(1):1–31.
- Betancourt, Roger R. and David A. Gautschi. 1996. An international comparison of the determinants of retail gross margins. *Empirica* 23(2):173–189.
- Bils, Mark and Peter J. Klenow. 2001. Quantifying quality growth. *American Economic Review* 91(4):1006–1030.
- Broda, Christian, Ephraim Leibtag, and David E. Weinstein. 2009. The role of prices in measuring the poor’s living standards. *Journal of Economic Perspectives* 23(2):77–97.
- Burnod, Guillaume and Alain Chenu. 2001. Employés qualifiés et non-qualifiés: Une proposition d’aménagement de la nomenclature des catégories socioprofessionnelles. *Travail et Emploi* 0(86):87–105.
- Carrasco, Marine and Jean-Pierre Florens. 2000. Generalization of GMM to a continuum of moment conditions. *Econometric Theory* 16(6):797–834.
- Chaney, Thomas, David Sraer, and David Thesmar. 2007. Collateral value and corporate investment: Evidence from the french real estate market. Working paper, INSEE-DESE G2007-08.
- Ciccone, Antonio and Robert E. Hall. 1996. Productivity and the density of economic activity. *American Economic Review* 86(1):54–70.

- Clauset, Aaron, Cosma Rohilla Shalizi, and M. E. J. Newman. 2009. Power-law distributions in empirical data. *SIAM Review* 51(4):661–703.
- Clerides, Sofronis, Saul Lach, and James R. Tybout. 1998. Is learning by exporting important? Micro-dynamic evidence from Colombia, Mexico, and Morocco. *Quarterly Journal of Economics* 113(3):903–947.
- Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63(2):723–742.
- Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2011. The costs of agglomeration: Land prices in French cities. Processed, University of Toronto.
- Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. Estimating agglomeration effects with history, geology, and worker fixed-effects. In Edward L. Glaeser (ed.) *Agglomeration Economics*. Chicago, IL: Chicago University Press, 15–65.
- De Loecker, Jan. 2011. Product differentiation, multi-product firms and estimating the impact of trade liberalization on productivity. *Econometrica* 79(5):1407–1451.
- Del Gatto, Massimo, Gianmarco I.P. Ottaviano, and Marcello Pagnini. 2008. Openness to trade and industry cost dispersion: Evidence from a panel of Italian firms. *Journal of Regional Science* 48(1):97–129.
- Duranton, Gilles and Diego Puga. 2001. Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review* 91(5):1454–1477.
- Duranton, Gilles and Diego Puga. 2004. Micro-foundations of urban agglomeration economies. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2063–2117.
- Ellison, Glenn and Edward L. Glaeser. 1999. The geographic concentration of industry: Does natural advantage explain agglomeration? *American Economic Review Papers and Proceedings* 89(2):311–316.
- Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review* 98(1):394–425.
- Fujita, Masahisa and Hideaki Ogawa. 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics* 12(2):161–196.
- Glaeser, Edward L. and David C. Maré. 2001. Cities and skills. *Journal of Labor Economics* 19(2):316–342.
- Gobillon, Laurent and Sébastien Roux. 2010. Quantile-based inference of parametric transformations between two distributions. Processed, CREST-INSEE.
- Handbury, Jessie H. and David E. Weinstein. 2010. Is new economic geography right? Evidence from price data. Processed, Columbia University.
- Head, Keith and Thierry Mayer. 2004. The empirics of agglomeration and trade. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2609–2669.
- Heckman, James J. and Bo E. Honore. 1990. The empirical content of the Roy model. *Econometrica* 58(5):1121–1149.

- Hellerstein, Judith K., David Neumark, and Kenneth R. Troske. 1999. Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations. *Journal of Labour Economics* 17(3):409–446.
- Henderson, J. Vernon. 1974. The sizes and types of cities. *American Economic Review* 64(4):640–656.
- Henderson, J. Vernon. 1997. Externalities and industrial development. *Journal of Urban Economics* 42(3):449–470.
- Holmes, Thomas J., Wen-Tai Hsu, and Sanghoon Lee. 2011. Plants, productivity, and market size, with head-to-head competition. Processed, University of Minnesota.
- Hopenhayn, Hugo. 1992. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica* 60(5):1127–1150.
- Jorgenson, Dale W., Mun S. Ho, and Kevin J. Stiroh. 2005. Growth of U.S. industries and investments in information technology and higher education. In Carol Corrado, John Haltiwanger, and Daniel Sichel (eds.) *Measuring Capital in the New Economy*. Chicago: University of Chicago Press, 260–304.
- Levinsohn, James and Amil Petrin. 2003. Estimating production functions using inputs to control for unobservables. *Review of Economic Studies* 70(2):317–342.
- Lileeva, Alla and Daniel Trefler. 2010. Improved market access to foreign markets raises plant-level productivity... for some plants. *Quarterly Journal of Economics* 125(3):1051–1099.
- Lucas, Robert E., Jr. and Esteban Rossi-Hansberg. 2002. On the internal structure of cities. *Econometrica* 70(4):1445–1476.
- Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan.
- Mayer, Thierry, Marc Melitz, and Gianmarco I. P. Ottaviano. 2011. Market size, competition, and the product mix of exporters. Discussion Paper 8349, Centre for Economic Policy Research.
- Melitz, Marc and Gianmarco I. P. Ottaviano. 2008. Market size, trade and productivity. *Review of Economic Studies* 75(1):295–316.
- Melitz, Marc J. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71(6):1695–1725.
- Melo, Patricia C., Daniel J. Graham, and Robert B. Noland. 2009. A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics* 39(3):332–342.
- Nocke, Volker. 2006. A gap for me: Entrepreneurs and entry. *Journal of the European Economic Association* 4(5):929–956.
- Olley, G. Steven and Ariel Pakes. 1996. The dynamics of productivity in the telecommunication equipment industry. *Econometrica* 64(6):1263–1297.
- Pavcnik, Nina. 2002. Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants. *Review of Economic Studies* 69(1):245–276.
- Roback, Jennifer. 1982. Wages, rents, and the quality of life. *Journal of Political Economy* 90(6):1257–1278.

- Rosenthal, Stuart S. and William Strange. 2004. Evidence on the nature and sources of agglomeration economies. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2119–2171.
- Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: Printed for W. Strahan, and T. Cadell.
- Sveikauskas, Leo. 1975. Productivity of cities. *Quarterly Journal of Economics* 89(3):393–413.
- Syversen, Chad. 2004. Market structure and productivity: A concrete example. *Journal of Political Economy* 112(6):1181–1222.

Appendix A. The TFP distribution is well approximated by a log-normal not by a Pareto

Recent models of firm selection often rely on the assumption that firms draw their values of TFP from a Pareto distribution. We have instead developed both our nested model of selection and agglomeration and our empirical approach without relying on any particular distribution. This appendix justifies the need for this generality by showing that the usual assumption that TFP follows a Pareto distribution, while analytically convenient, is unrealistic. If anything, the empirical TFP distribution is well approximated by a log-normal.

To show this, we fit the empirical TFP distribution with a mixture of a log-normal distribution (with weight μ) and a Pareto distribution (with weight $1 - \mu$). This mixture has the probability density function

$$f_M(x) = \mu f_N(x) + (1 - \mu) f_P(x) ,$$

where

$$f_N(x) = \frac{1}{x\sqrt{2\pi v}} e^{-\frac{(\ln(x)-m)^2}{2v}}$$

denotes the density of a log-normal distribution with mean m and variance v , and

$$f_P(x) = \begin{cases} 0 & \text{for } x < b , \\ zb^z x^{-z-1} & \text{for } x \geq b , \end{cases}$$

denotes the density of a Pareto distribution with minimum value b and shape parameter z .²⁸

The approach used to approximate the empirical TFP distribution with the mixed distribution is similar to the one used in the main text to approximate the TFP distribution in denser areas by shifting, dilating, and truncating the distribution in less dense areas. The set of parameters we must now estimate is $\zeta = (\mu, m, v, b, z)$. The main difference is that we base our estimation on the cumulative density function of the mixture F_M to avoid convergence problems. These convergence problems are caused by the extremely high values that quantiles can take at high ranks with a Pareto distribution, a feature that is not present in the empirical distribution. Focusing on the

²⁸There is a literature that tests whether a pure Pareto distribution provides a good fit for a number of empirical distributions (see, in particular, Clauset, Shalizi, and Newman, 2009). We are instead checking what mixture of a log-normal distribution and a Pareto distribution provides the best fit.

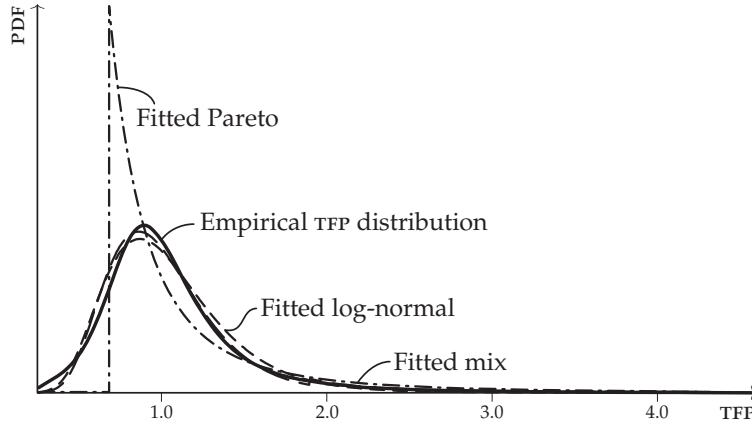


Figure A.1: Empirical TFP distribution, and fitted Pareto, log-normal, and mixed distributions

cumulative facilitates convergence for the Pareto component of the mixture. The estimator we use is

$$\hat{\zeta} = \arg \min_{\zeta} C(\zeta), \quad \text{where} \quad C(\zeta) = \frac{1}{E} \sum_{k=1, \dots, E} [k/E - F_M(x(k))]^2,$$

where $k = 1, \dots, E$ indexes the E establishments or observations of TFP.

Using the empirical TFP distribution from our baseline results (OLS estimates of TFP for all sectors combined), we find $\hat{\zeta} = (\hat{\mu}, \hat{m}, \hat{\sigma}, \hat{b}, \hat{z}) = (0.95, -0.05, 0.32, 1.90, 1.89)$. The key parameter is $\hat{\mu} = 0.95$, i.e., the empirical TFP distribution is best approximated by a mixture that is 95% log-normal and 5% Pareto. Another interesting finding is that $\hat{b} = 1.90$, i.e., the Pareto component of the mixture is only used to improve the fit starting from $x = 1.90$. As illustrated in figure A.1, this is already very high in the upper tail of the empirical TFP distribution (one-and-a-half standard deviations above the mean). In addition to the empirical TFP distribution and the fitted mix of Pareto and log-normal, the figure also plots two restricted versions of the fitted distribution.²⁹ We first re-estimate $\hat{\zeta}$ with the restriction $\mu = 1$, which forces the fitted distribution to be 100% log-normal. The mean and the variance increase slightly, to $\hat{m} = -0.02$ and $\hat{\sigma} = 0.35$, relative to the log-normal component of the mixed distribution fitted before. This partly offsets the loss of the Pareto component to help fit the very upper tail, at the expense of losing some accuracy in the fit for the rest of the distribution. We then impose the opposite restriction $\mu = 0$, which forces the empirical TFP distribution to be 100% Pareto. Looking at the fitted Pareto in the figure makes it clear how far its shape is from the empirical TFP distribution. Parameters change substantially to $\hat{b} = 0.68$ and $\hat{z} = 2.14$, as the estimation now struggles to fit the bottom and middle of the distribution using a Pareto alone.

There are two reasons, besides analytical convenience, why it is often assumed that TFP follows a Pareto distribution. First, instead of looking at the TFP distribution, some studies look at the

²⁹Note that figure A.1 plots TFP, not log TFP as other figures in the paper, because papers using the Pareto assumption make this about productivity in levels and not in logs. A Pareto distribution for TFP implies an exponential distribution for log TFP. Trying to fit a mixture of normal and exponential on log TFP (as opposed to log-normal and Pareto on TFP) yields similar results. In addition, using our Olley-Pakes or Levinsohn-Petrin TFP estimates instead of OLS TFP estimates yields similar estimates for $\hat{\zeta}$.

size distribution of firms by employment and use models where there is a one-to-one mapping between TFP and employment. Second, other studies look at the TFP distribution, but focus on the upper tail only. However, while cutting everything below the mode of a unimodal distribution can make it visually similar to a Pareto, it is not necessarily so. To assess this more formally, we next extend our procedure to truncate the empirical TFP distribution at its mode and then approximate the upper tail with a mixture of a log-normal left-truncated at its mode and a Pareto. Even in this case where we ignore everything to the left of the peak of the empirical TFP distribution and focus only on the upper tail of the distribution, we find that this upper tail is best approximated by a mixture that is 91% a log-normal truncated at its mode and 9% Pareto.

To summarize, the empirical TFP distribution is well approximated by a log-normal distribution, although the very upper tail of the distribution is slightly fatter than one would expect from a log-normal distribution.

Appendix B. Proof of Proposition 1

Consider any two areas i and j such that $i < j$ (and thus $N_i > N_j$). The dilation factor is D_i in cities i and D_j in city j while the extent of the right shift is A_i in city i and A_j in city j . If $0 \leq \delta < 1$, by equation (11), $D_i > D_j$ and, by equation (7), $A_i > A_j$. If instead $\delta = 1$, by the same two equations, $D_i = D_j$ and $A_i = A_j$. Turning to selection, the proportion of truncated values of \tilde{F} is S_i in city i and S_j in city j . The free entry condition (6) for cities i and j can be rewritten:

$$\frac{N_i}{4\gamma} \int_0^{\bar{h}_i} (\bar{h}_i - h)^2 g(h) dh + \frac{N_j}{4\gamma} \int_0^{\bar{h}_j/\tau} (\bar{h}_j - \tau h)^2 g(h) dh + \sum_{k \neq i, k \neq j} \frac{N_k}{4\gamma} \int_0^{\bar{h}_k/\tau} (\bar{h}_k - \tau h)^2 g(h) dh = s, \quad (\text{B.1})$$

$$\frac{N_j}{4\gamma} \int_0^{\bar{h}_j} (\bar{h}_j - h)^2 g(h) dh + \frac{N_i}{4\gamma} \int_0^{\bar{h}_i/\tau} (\bar{h}_i - \tau h)^2 g(h) dh + \sum_{k \neq i, k \neq j} \frac{N_k}{4\gamma} \int_0^{\bar{h}_k/\tau} (\bar{h}_k - \tau h)^2 g(h) dh = s. \quad (\text{B.2})$$

Subtracting equation (B.2) from (B.1) and simplifying yields:

$$N_i v(\bar{h}_i, \tau) = N_j v(\bar{h}_j, \tau). \quad (\text{B.3})$$

where

$$v(z, \tau) \equiv \int_0^z (z - h)^2 g(h) dh - \int_0^{z/\tau} (z - \tau h)^2 g(h) dh. \quad (\text{B.4})$$

It follows from (B.3) and $N_i > N_j$ that

$$v(\bar{h}_i, \tau) < v(\bar{h}_j, \tau). \quad (\text{B.5})$$

Differentiating (B.4) with respect to z yields:

$$\begin{aligned} \frac{\partial v(z, \tau)}{\partial z} &= 2 \left[\int_0^z (z - h) g(h) dh - \int_0^{z/\tau} (z - \tau h) g(h) dh \right] \\ &= 2 \left[(\tau - 1) \int_0^{z/\tau} h g(h) dh + \int_{z/\tau}^z (z - h) g(h) dh \right]. \end{aligned} \quad (\text{B.6})$$

If $1 < \tau < \infty$, then $\partial v(z, \tau) / \partial z > 0$, and thus, by equation (B.5), $\bar{h}_i < \bar{h}_j$. Hence, by equation (9), $S_i > S_j$. If $\tau = 1$, then by equation (B.6), $\partial v(z, \tau) / \partial z = 0$, and thus $\bar{h}_i = \bar{h}_j$ and $S_i = S_j$. \square

Appendix C. Proof of Lemma 1

Consider first the case $S_i > S_j$. We apply the change of variables $\phi \rightarrow \frac{\phi - A}{D}$, which turns the expression for F_j that follows from equation (13) into

$$F_j\left(\frac{\phi - A}{D}\right) = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_j}{1 - S_j}\right\}.$$

Dividing by $1 - S$ and adding $\frac{-S}{1-S}$ to all terms in this equation yields

$$\frac{F_j\left(\frac{\phi - A}{D}\right) - S}{1 - S} = \max\left\{\frac{-S}{1 - S}, \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_i}{1 - S_i}\right\}.$$

Since, with $S_i > S_j$, $S > 0$, we have $\frac{-S}{1-S} < 0$, and we obtain

$$\max\left\{0, \frac{F_j\left(\frac{\phi - A}{D}\right) - S}{1 - S}\right\} = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_i}{1 - S_i}\right\} = F_i(\phi).$$

Consider now the case $S_i < S_j$. We apply the change of variables $\phi \rightarrow D\phi + A$, which turns equation (13) into

$$F_i(D\phi + A) = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_j}{D_j}\right) - S_i}{1 - S_i}\right\}.$$

Dividing by $1 - \frac{-S}{1-S}$ and adding S to all terms in this equation yields

$$\frac{F_i(D\phi + A) - \frac{-S}{1-S}}{1 - \frac{-S}{1-S}} = \max\left\{S, \frac{\tilde{F}\left(\frac{\phi - A_j}{D_j}\right) - S_j}{1 - S_j}\right\}.$$

Since, with $S_i < S_j$, $S < 0$, we finally obtain

$$\max\left\{0, \frac{F_i(D\phi + A) - \frac{-S}{1-S}}{1 - \frac{-S}{1-S}}\right\} = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_j}{D_j}\right) - S_j}{1 - S_j}\right\} = F_j(\phi).$$

□

Appendix D. The consequences of unobserved prices when $S \neq 0$

We now explore the consequences for our methodology of not observing prices when, contrary to our empirical findings, $S \neq 0$. Consider first the case where, as in our model when markets are not closely integrated, $S > 0$. Expressed in terms of the model, the inability to observe prices implies that we do not measure ϕ , as given by equation (12), but instead

$$\psi = \ln\left(\frac{pQ}{l}\right) = \ln(p) + A_i - D_i \ln(h) = \ln(p) + \phi.$$

Thus, by not taking prices out, we are shifting log productivity by the value of log prices, $\ln(p)$.

The problem is that, if $S > 0$, log prices are systematically related both to city size (through \bar{h}) and to individual productivity (through h) since, by equation (5), prices are given by $p = \frac{1}{2}(\tau h + \bar{h})$. In terms of the relationship with city size,

$$\frac{\partial \ln(p)}{\partial \bar{h}} = \frac{1}{\tau h + \bar{h}} > 0 .$$

If \bar{h} differs across cities, then by looking at ψ instead of ϕ we are obtaining a biased estimate of log productivity for every h , but the bias is larger (more positive) in smaller cities, where \bar{h} is then larger. Hence, when $S > 0$, one consequence of not observing prices is that we will underestimate A , the parameter capturing the common shift in the log productivity distribution of large cities relative to small cities. In terms of the relationship with individual productivity,

$$\frac{\partial^2 \ln(p)}{\partial \bar{h} \partial h} = -\frac{\tau}{(\tau h + \bar{h})^2} < 0 .$$

Thus, if \bar{h} differs across cities, the problem of underestimating log productivity in large versus small cities is greater for the most productive firms. Hence, when $S > 0$, another consequence of not observing prices is that we will underestimate D , the parameter capturing to what extent more productive firms get an extra productivity boost from locating in large cities.

We have thus shown that, if $S > 0$, then by not observing prices we would underestimate both A and D . If instead $S < 0$, the argument is reversed and we will overestimate both A and D . Finally, if $S = 0$, then \bar{h} does not vary with city size, and the estimates of A and D are unbiased. Recall that in our empirical we find \hat{S} close to 0 for all sectors combined and for nearly all individual sectors. Note also that not observing prices does not affect the estimation of S , since this is defined as the share of establishments in the small city distribution that are truncated out of the large city distribution. Thus, our finding that \hat{S} is close to 0 also implies that estimating TFP through value added does not bias our estimates of A and D .