University of Toronto
Department of Economics

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

By Pierre-Philippe Combes, Gilles Duranton, Diego Puga and Sebastien Roux

March 05, 2009

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

Pierre-Philippe Combes*†
University of Aix-Marseille and *CEPR*

Gilles Duranton*‡
University of Toronto and *CEPR*

Laurent Gobillon*§
Institut National d'Etudes Démographiques, *PSE-INRA*, and *CREST*

Diego Puga*§
IMDEA, *Universidad Carlos III* and *CEPR*

Sébastien Roux*‖
CREST - INSEE

February 2009

ABSTRACT: Firms are more productive on average in larger cities. Two explanations have been offered: agglomeration economies (larger cities promote interactions that increase productivity) and firm selection (larger cities toughen competition allowing only the most productive to survive). To distinguish between them, we nest a generalised version of a seminal 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. We assess the relative importance of agglomeration and firm selection using French establishment-level data and a new quantile approach. Spatial productivity differences in France are mostly explained by agglomeration.

Key words: agglomeration, firm selection, productivity, cities

JEL classification: C52, R12, D24

†GREQAM, 2 Rue de la Charité, 13236 Marseille cedex 02, France (e-mail: ppcombes@univmed.fr; website: http://www.vcharite.univ-mrs.fr/pp/combes/).

‡Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario M5S 3G7, Canada (e-mail: gilles.duranton@utoronto.ca; website: http://individual.utoronto.ca/gilles/default.html).

§Institut National d'Etudes Démographiques, 133 Boulevard Davout, 75980 Paris cedex 20, France (e-mail: laurent.gobillon@ined.fr; website: http://laurent.gobillon.free.fr/).

§Madrid Institute for Advanced Studies (IMDEA) Social Sciences, Antiguo pabellón central del Hospital de Cantoblanco, Carretera de Colmenar Viejo km. 14, 28049 Madrid, Spain (e-mail: diego.puga@imdea.org; website: http://diegopuga.org).

‖Centre de Recherche en Économie et Statistique (CREST), 15 Boulevard Gabriel Péri, 92245 Malakoff cedex, France (e-mail: sebastien.roux@ensae.fr).

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 magnitude of this effect range between a 2 and 7 percent productivity increase from a doubling of city size for a large range of city sizes, depending on the sector and details of the estimation procedure.

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 (see Duranton and Puga, 2004, for a review). All these mechanisms share a common prediction: the concentration of firms and workers in space makes them more productive.

More recently, an alternative explanation has been offered based on 'firm selection'. The argument builds on work by Melitz (2003), who introduces product differentiation and international or inter-regional trade in the framework of industry dynamics of Hopenhayn (1992). Melitz and Ottaviano (2008) incorporate endogenous price-cost mark-ups in this framework and show that larger markets attract more firms, which makes competition tougher.[1] In turn, this leads less productive firms to exit. This suggests that the higher average productivity of firms and workers in larger cities could instead result from a stronger Darwinian selection of firms.

Our main objective in this paper is to distinguish between agglomeration and firm selection in explaining why average productivity is higher in larger cities. To do so, our first step 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 Rossi-Hansberg (2002). This nested model allows us to parameterise the relative importance of agglomeration and selection. The main prediction of our model is that, while selection and agglomeration effects both make average firm log productivity higher in larger cities, they have different predictions for how the shape of the log productivity distribution varies with city size. In particular, stronger selection effects in larger cities should lead to a greater left truncation of the distribution of firm log productivities in larger cities, as the least productive firms exit. Stronger agglomeration effects in larger cities should lead instead to a greater rightwards shift of the distribution of firm log productivities in larger cities, as agglomeration effects make all firms more productive. 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 productivities in larger cities.

We then use this predictions to assess the relative importance of agglomeration and firm selection for different sectors using data for all French firms. Our structural estimation is in two steps.

---

[1] Bernard, Eaton, Jensen, and Kortum (2003) also develop a model with heterogenous firm productivity levels and endogenous mark-ups but, unlike in Melitz and Ottaviano (2008), these mark-ups are not affected by market size. In Nocke (2006), more able entrepreneurs sort into larger markets because competition there is more intense.

We first estimate total factor productivity at the establishment level. Next, we develop a new quantile approach to compare the distribution of establishment log productivities for each sector across metropolitan areas of different sizes. As stipulated by the model, we estimate the extent to which the log productivity distribution in large cities is left-truncated (evidence of differences in selection effects) or dilated and right-shifted (evidence of differences in agglomeration effects) compared to the log productivity distribution in small cities.

This empirical approach offers a number of benefits. First, it allows both agglomeration economies and firm selection 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 and on agglomeration economies raising everyone's productivity — possibly to different extents. Third, we do not rely on particular distributional assumptions of firms' 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 greater 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.

Our main finding is that productivity differences between French metropolitan areas are explained mostly by agglomeration. On the other hand, we find no systematic evidence of stronger selection in larger cities. We begin with the simplest characterisation of agglomeration economies: a common upwards shift in log productivity. This shift alone is able to explain most of the differences in the log productivity distribution between cities of different sizes, and corresponds to a productivity gain across the board of about 12 percent for establishments in metropolitan areas with population above 200,000 relative to establishments located elsewhere. Even with this simple characterisation of agglomeration economies, there are just no sizeable differences in left truncation across cities of different sizes. We then also allow for the possibility that more productive establishments are better able to reap the benefits from agglomeration, which dilates upwards the log productivity distribution. This additional consequence of agglomeration is also supported by the data. While the average productivity gain is now about 9 percent , establishments at the bottom quartile of the log productivity distribution are only 5 percent more productive in metropolitan areas with population above 200,000 than elsewhere whereas establishments at the top quartile are about 14 percent more productive in larger cities.

While our results about agglomeration and selection apply broadly to manufacturing and business services as well as to most particular sectors, we can find some exceptions within sufficiently detailed sectoral classifications. A few sectors do not seem to benefit from stronger agglomeration economies in large metropolitan areas. A few sectors also exhibit stronger selection in large metropolitan areas. However, such exceptions play almost no role in explaining differences across cities in the aggregate log productivity distribution. Finally, none of our results appears sensitive to our choice of estimation technique for productivity nor to the sample of establishments.

Our paper is 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. It is also related to the pioneering work of Syverson (2004) who examines the effect of market size on firm selection in the ready-made concrete sector and the emerging literature that follows (Del Gatto, Mion, and Ottaviano, 2006, 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 on a model incorporating selection alone. Considering agglomeration and selection simultaneously 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 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.[2]

The rest of this paper is organised as follows. The next section proposes a generalisation of Melitz and Ottaviano (2008) and combines it with an agglomeration model. Section 3 describes our econometric approach. Section 4 discusses the data and the details of our empirical implementation. The baseline results are then presented in Section 5. Section 6 introduces a more general version of our theoretical model and econometric approach, and Section 7 presents our corresponding main results. Finally, Section 8 discusses some additional issues, and Section 9 concludes.

2. A nested model of selection and agglomeration

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 in the spirit of Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002).

An individual consumer's utility is given by

$$U = q^0 + \alpha \int_{i \in \Omega} q^i \mathrm{d}i - \frac{1}{2}\gamma \int_{i \in \Omega} \left(q^i\right)^2 \mathrm{d}i - \frac{1}{2}\eta \left(\int_{i \in \Omega} q^i \mathrm{d}i \right)^2 , \qquad (1)$$

---

[2]There 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) conclude at the predominance of self-selection by observing that exporting firms have better pre-determined characteristics. More recent work by Lileeva and Trefler (2007) 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 (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 identifies 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 postpone further discussion of how our results fit with the implications from this trade literature to the concluding section.

where q^0 denotes the individual's consumption of a homogenous numéraire good, and q^i her consumption of variety i of a set Ω of differentiated varieties. The three positive demand parameters α, γ, and η are such that a higher α and a lower η increase demand for differentiated varieties relative to the numéraire, while a higher γ reflects more product differentiation between varieties.[3]

Utility maximisation subject to the budget constraint yields the following inverse demand for differentiated variety i by an individual consumer:

$$p^i = \alpha - \gamma q^i - \eta \int_{j \in \Omega} q^j \mathrm{d}j \, , \tag{2}$$

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

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

The price threshold, $\bar{h}$, in equation (3) follows immediately from the restriction $q^i \geqslant 0$. 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 when we consider more than one location. This implies that the cost to firms of hiring one unit of labour is always unity.[4]

Differentiated varieties are produced under monopolistic competition. By incurring a sunk entry cost s, a firm is able to develop a new variety that can be produced 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 unit labour requirement h differs across firms and for each of them it is randomly drawn from a distribution with known probability density function $g(h)$ and cumulative $G(h)$. Melitz and Ottaviano (2008) derive most of their results under the assumption that $g(h)$ is a Pareto distribution. By contrast, we do not adopt any particular distribution for $g(h)$. To simplify the derivation of the results, we only require $G(.)$ to be continuous and differentiable. Firms with a marginal cost higher than the price at which consumer demand becomes zero are unable to cover their marginal cost and exit. The set of goods varieties that end up being produced in equilibrium is therefore $\bar{\Omega} = \{i \in \Omega \mid h \leqslant \bar{h}\}$.

Since all varieties enter symmetrically into utility, we can index firms by their unit labour requirement h instead of the specific variety i they produce. Re-writing the individual consumer demand of (3) in terms of $\bar{h}$ and multiplying this by the mass of consumers C yields the following

---

[3]The specification in (1) is often referred to as the quadratic utility model of horizontal product differentiation. It has been used in industrial organisation by, for instance, Dixit (1979) and Vives (1990) and has become popular in location modelling following Ottaviano, Tabuchi, and Thisse (2002).

[4]The 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.

expression for the total sales of an individual firm:

$$Q(h) = Cq(h) = \begin{cases} \frac{C}{\gamma}[\bar{h} - p(h)] & \text{if } p(h) \leqslant \bar{h} \,, \\ 0 & \text{if } p(h) > \bar{h} \,. \end{cases} \quad (4)$$

Given that the entry cost is sunk when firms draw their value of h, active firms set prices to maximise operational profits given by

$$\pi(h) = [p(h) - h]Q(h) \,. \quad (5)$$

Maximising $\pi(h)$ in (5) subject to (4) yields the optimal pricing rule

$$p(h) = \frac{1}{2}(h + \bar{h}) \,. \quad (6)$$

Substituting (4) and (6) into (5) we obtain equilibrium operational profits:

$$\pi(h) = \frac{C}{4\gamma}(\bar{h} - h)^2 \,. \quad (7)$$

Entry into the monopolistically competitive industry takes place until *ex-ante* expected profits are driven to zero. The operational profits expected prior to entry must therefore be exactly offset by the sunk entry cost:

$$\frac{C}{4\gamma} \int_0^{\bar{h}} (\bar{h} - h)^2 g(h)\mathrm{d}h = s \,. \quad (8)$$

The free-entry condition (8) implicitly defines the marginal cost cutoff $\bar{h}$ as a function of the distribution $g(h)$, the sunk entry cost s, the mass of consumers C, and the degree of product differentiation parameter γ. We note that while we rely on the framework developed by Melitz and Ottaviano (2008), the existence of a marginal cost cutoff should be a general property of a whole class of market selection models.

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. More specifically, when interacting with W other workers, the effective units of labour supplied by an individual worker during their unit working time is $a(W)$, where $a(0) = 1$, $a' > 0$ and $a'' < 0$. 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) — we introduce a discrete version of their spatial decay for interactions below. Duranton and Puga (2004) review micro-foundations of numerous alternative agglomeration mechanisms, which also result in a reduced form like $a(W)$. We assume that such interactions benefit workers across occupations, i.e., regardless of whether they produce any particular variety of the differentiated

good or the numéraire good (a simplifying assumption that we relax below).[5] This, given the unit payment per effective unit of labour supplied, implies that the total labour income of each worker in any occupation is $a(W)$.

A firm with unit labour requirement h hires $l(h) = Q(h)h/a(W)$ workers at a total cost of $a(W)l(h) = Q(h)h$. The natural logarithm of the firm's productivity is then given by

$$\phi = \ln\left(\frac{Q}{l}\right) = \ln\left[a(W)\right] - \ln(h) \ . \tag{9}$$

The probability density function of firms' log productivities is then

$$f(\phi) = \begin{cases} 0 & \text{for } \phi < A - \ln(\bar{h}) \ , \\ \dfrac{e^{A-\phi}g\left(e^{A-\phi}\right)}{G(\bar{h})} & \text{for } \phi \geqslant A - \ln(\bar{h}) \ , \end{cases} \tag{10}$$

where

$$A \equiv \ln\left[a(W)\right] \ . \tag{11}$$

The numerator of $f(\phi)$, $e^{A-\phi}g\left(e^{A-\phi}\right)$ follows from using equation (9) and the change of variables theorem, while the denominator $G(\bar{h})$ takes care of the fact that firms with a unit labour requirement above $\bar{h}$ exit. The model can now be solved sequentially by first using the free entry condition of equation (8) to solve for the equilibrium cut-off unit labour requirement $\bar{h}$. We can then substitute $\bar{h}$ into (10) to obtain the equilibrium distribution of firm productivities. Finally, equation (6) gives prices and the definition of $\bar{h}$ in (3) allows us to compute the mass of varieties produced, ω.

To understand how selection and agglomeration forces contribute to determining the distribution of firms' log productivities we must consider what is the relevant mass C of consumers that each firm sells to and what is the relevant mass W of people that each worker interacts with. Before deriving our formal results in a general setting below, let us first develop a simpler example that illustrates the key properties of our framework with particular clarity.

An illustrative example

Consider two polar possibilities for both demand and interactions in an economy with two cities. In terms of demand, at one extreme we can think of firms selling only to consumers in their city and thus competing with other local firms only (*local product-market competition*). At the other extreme, firms can sell with equal ease to consumers anywhere and thus compete with firms everywhere (*global product-market competition*). In terms of interactions, at one extreme we can think of workers interacting exclusively with other workers living in the same city (*local interactions*). At the other

---

[5]More realistically, we would expect the benefits of agglomeration to depend not just on the interactions between workers but also on the characteristics of firms. In particular, more productive firms (i.e., those with a lower h) may be more able to reap the benefits of agglomeration. We explicitly allow for this possibility in Section 6, where we extend both our model and empirical approach. Regarding the assumption that the benefits of agglomeration economies percolate to the numéraire sector, this ensures that such benefits get reflected in individual worker earnings so that these increase with city size, consistent with the empirical evidence (see, e.g., Glaeser and Maré, 2001, Wheaton and Lewis, 2002, Combes, Duranton, Gobillon, and Roux, 2009). If we re-wrote the model so that, counterfactually, earnings per worker not only did not increase but decreased with city size, then selection forces could be much weakened or even reversed.
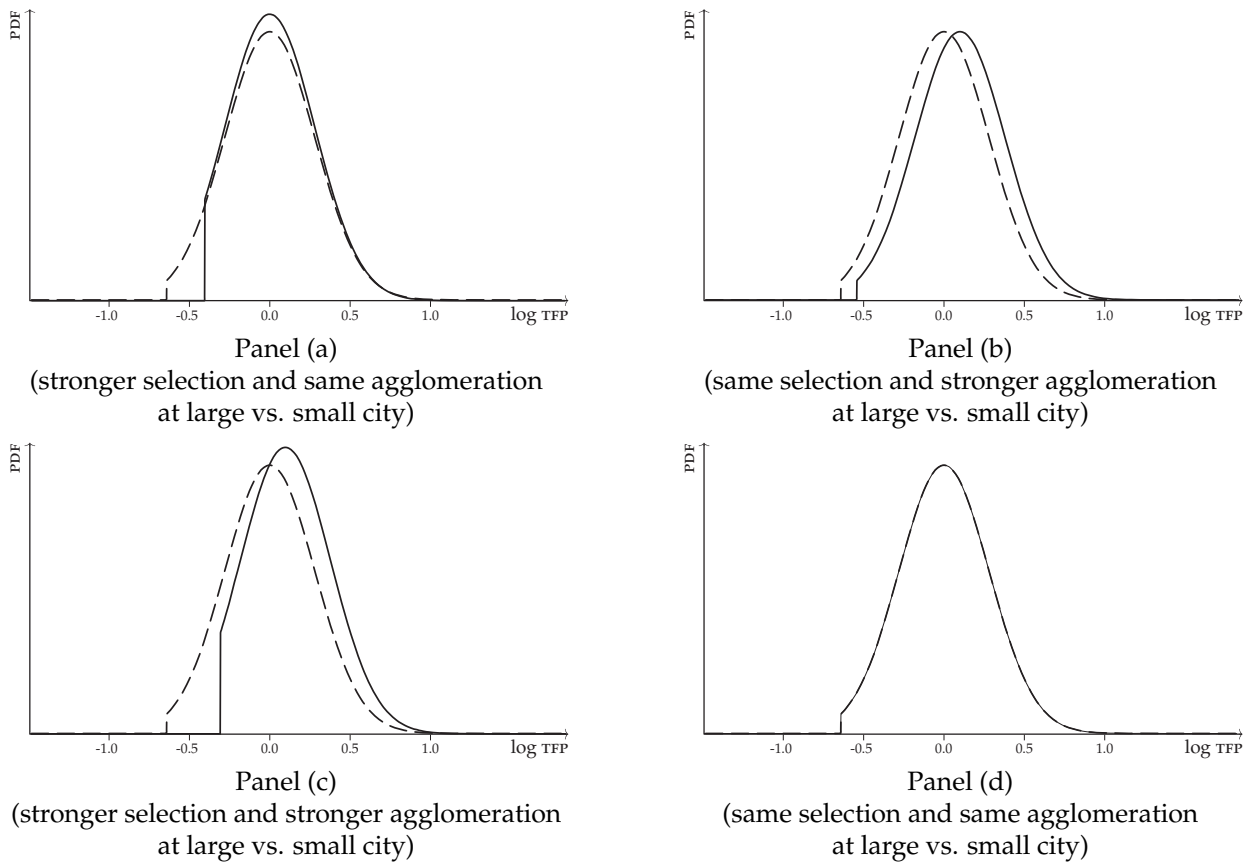
Figure 1: Log productivity distributions in large (solid) and small cities (dashed)

extreme, workers can interact with equal ease with workers living anywhere (*global interactions*). The combination of these possibilities gives us four cases.[6] We now compare in each of the four cases the distribution of firms' log productivities across two cities of different population size (a large city indexed 1 and a small city indexed 2).

Case 1 (*local product-market competition and global interactions*). Panel (a) in Figure 1 plots the distribution of firms' log productivities in a city with a large population (continuous line) and in a city with a small population (dashed line) in the case where firms only sell in their local city and workers enjoy interactions with the same intensity with workers from everywhere (i.e., $C_1 > C_2$ and $W_1 = W_2$). Compared with the distribution in the small city, the large-city distribution is left-truncated as a consequence of firm selection. This left truncation implies that the peak of the large city distribution is higher than that of the small city distribution and that the two peaks occur at the same level of productivity. To understand this greater truncation in the large city, note that 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. Formally, using equations (4) and (6), total sales for an individual firm can be expressed as $\frac{C}{2\gamma}(\bar{h} - h)$. Hence, for a given measure of firms ω, that is, for a given $\bar{h}$,

---

[6]We later derive our formal results in a general setting with multiple cities where firms can sell to consumers everywhere, but there are potentially some additional costs incurred when they sell outside their city, and workers interact with other workers everywhere, but interactions with workers in other cities are potentially weaker.

sales increase proportionately with the population of consumers C. However, the larger individual firm sales associated with a larger C make further entry profitable and, by equation (8), they must be offset by a lower $\bar{h}$ to restore zero *ex-ante* expected profits.[7]

To understand how firms in different ranges of the productivity distribution are affected by city size and $\bar{h}$, note that, from (4) and (6), the price elasticity of demand faced at equilibrium by a firm with unit labour requirement h can be written as follows:

$$\epsilon(h) \equiv -\frac{p(h)}{Q(h)}\frac{\mathrm{d}Q(h)}{\mathrm{d}p(h)} = \frac{\bar{h}+h}{\bar{h}-h}. \tag{12}$$

Demand becomes more price-elastic as h increases or as $\bar{h}$ decreases. Thus, each firm in the large city (where $\bar{h}$ is lower) faces a more elastic demand, and hence charges a lower markup, $p - h = (\bar{h} - h)/2$, than a firm with the same h in the small city. The combination of more consumers, further entry, and the ensuing lower markups in the large city affects firms' sales differently depending on their h.[8] In the large city, firms with high productivity, and hence high markups, enjoy smaller profit margins but larger sales than their small-city counterparts. Low productivity firms, however, have both smaller profit margins and smaller sales in the large city than in the small city. In short, product market competition is tougher in a large city than in a small city, and this affects firms with low productivity and hence low price-cost margins the most. 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. It is this exit at the low-productivity end that leads to the large-city log productivity distribution being a left-truncated version of the small-city distribution (see Lemma 1 below for a formal proof).

Case 2 (*global product-market competition and local interactions*). Panel (b) in Figure 1 plots the distribution of firms' log productivities in a city with a large population (continuous line) and in a city with a small population (dashed line) in the case where every firm competes with the same intensity with firms from everywhere and workers only interact with workers in their city (i.e., $C_1 = C_2$ and $W_1 > W_2$).[9] Compared with the distribution in the small city, the large-city distribution is right-shifted. Since interactions are local, workers in the larger city benefit from

---

[7]The reason why a larger measure of firms leads to a lower $\bar{h}$ is the following. By (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 varieties drop to zero. In turn, by (6), this lower $\bar{h}$ induces firms to lower their prices which, by (3), further reduces $\bar{h}$.

[8]This is best seen by considering the effect on a firm's sales $Q(h)$ of a small increase in C. From (4) and (6), $\frac{\mathrm{d}Q(h)}{\mathrm{d}C} = \frac{1}{2\gamma}\left[\bar{h} - h + C\frac{\mathrm{d}\bar{h}}{\mathrm{d}C}\right]$. From the free entry condition of (8), $\frac{\mathrm{d}\bar{h}}{\mathrm{d}C} = -2\gamma s/[C^2\int_0^{\bar{h}}(\bar{h} - h)g(h)\mathrm{d}h]$. It follows that $\frac{\mathrm{d}Q(h)}{\mathrm{d}C} > 0$ if and only if $(\bar{h} - h)\frac{C}{2\gamma}\int_0^{\bar{h}}(\bar{h} - h)g(h)\mathrm{d}h > s$. The expression on the left-hand side of this inequality is twice the firm's markup times *ex-ante* expected sales. Since there are zero expected *ex-ante* profits, for firms near the top of the productivity distribution (those with the lowest values of h and thus highest markup), twice their *ex-post* markup times *ex-ante* expected sales must be higher than the sunk entry cost, and this inequality holds, so their sales increase as city size increases. For firms near the bottom of the productivity distribution (those with a value of h close to the cutoff $\bar{h}$), the inequality fails to hold, so their sales fall as city size increases.

[9]To facilitate visual comparisons, we re-scale the combined size of the large and small cities from panel to panel to keep the sets C and W for the small city constant across cases, thus making the distribution of firms' log productivities in the small city identical in all four panels. This is done for the purpose of plotting the graph only, and does not change the qualitative comparison between the small-city and large-city distributions. These graphs are drawn using a log-normal distribution, but recall that our analytical results are distribution-independent. We use a log-normal distribution for the graphs because it matches well the empirically-observed distributions presented later in the paper.

being exposed to a wider range of ideas than workers in the small city and this makes them more productive. As a result, all large-city firms achieve higher log productivity than their small-city counterparts (i.e., log productivity ϕ is higher in the large city for every h). Since product-market competition is global, all firms can sell to consumers everywhere and this eliminates differences between cities in the strength of the firm selection mechanism. Hence, the log productivity cut-off $\ln[a(W)] - \ln(\bar{h})$ simply moves rightwards to the same extent as the rest of the log productivity distribution. Consequently the large-city log productivity distribution is simply a right-shifted version of the small-city distribution (again, see Lemma 1 below for a formal proof). Thus, agglomeration acts like the tide that lifts all boats.

Case 3 (*local product-market competition and local interactions*)**.** Panel (c) in Figure 1 plots the distribution of firms' log productivities in a city with a large population (continuous line) and in a city with a small population (dashed line) in the case where firms only sell in their local market and workers only interact with workers in their city (i.e., $C_1 > C_2$ and $W_1 > W_2$). Compared with the distribution in the small city, the large-city distribution is both left-truncated and right-shifted (again, see Lemma 1 below for a formal proof). With local product-market competition, large-city markups are lower and this left-truncates the distribution of firms' log productivities to exactly the same extent as under case 1. With local interactions, large-city workers are more productive and this right-shifts the distribution of firms' log productivities (truncated by firm selection) to exactly the same extent as under case 2.[10]

Case 4 (*global product-market competition and global interactions*)**.** When every firm competes with the same intensity with firms from everywhere and every worker enjoys interactions with the same intensity with workers from everywhere (i.e., $C_1 = C_2$ and $W_1 = W_2$), the distribution of firms' log productivities in a city with a large population is exactly the same as in a city with a small population. Panel (d) in Figure 1 plots the probability density function of the distribution of firms' log productivities, $f(\phi)$, in this final case where it is independent of city size. Note that the fact that the distribution of firms' log productivities does not depend on city size in this case does not imply that there are no selection or agglomeration effects. It simply implies that selection and agglomeration effects are equally strong everywhere. If there were no selection or agglomeration effects, the distribution of firm productivities would be given by $f(\phi) = e^{-\phi} g(e^{-\phi})$. Relative to this underlying distribution, the actual distribution of firms' log productivities in both cities is both left-truncated and right-shifted.

---

[10]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), which eliminates income effects in the market for differentiated varieties. 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 log productivity advantages of agglomeration would translate into a larger market for differentiated varieties 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. More complicated interactions between selection and agglomeration mechanisms would also appear if the benefits from agglomeration for each worker varied depending on which firm they were working for. See Section 6.

Parameterising the strength of selection and agglomeration

For expositional clarity, we have so far focused on the polar possibilities of either local or global product-market competition and either local or global interactions. We now generalise our analysis to also consider intermediate cases for both. We do so by parameterising both the spatial decay of product market competition, creating differences in firm selection across cities, and the spatial decay of interactions, creating differences in agglomeration economies across cities. As before, we compare the distribution of log productivities across cities of different size. Suppose we have I cities. Let us denote the population of city i by N_i and order cities from largest to smallest in terms of population: $N_1 > N_2 > \cdots > N_{I-1} > N_I$.

In the case of product-market competition, we can introduce additional costs for selling differentiated varieties in a different city. Suppose that markets are segmented and that selling outside the city where a firm is located involves iceberg trade costs so that τ (> 1) units need to be shipped for one unit to arrive at destination. Since firms now potentially sell in all cities, the free entry condition of equation (8) becomes

$$\frac{N_i}{4\gamma} \int_0^{\bar{h}_i} (\bar{h}_i - h)^2 g(h)\mathrm{d}h + \sum_{j \neq i} \frac{N_j}{4\gamma} \int_0^{\bar{h}_j/\tau} (\bar{h}_j - \tau h)^2 g(h)\mathrm{d}h = s \, , \tag{13}$$

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 only city i firms with marginal costs $h < \bar{h}_j/\tau$ 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. Note also that the cases of purely local or purely global product-market competition discussed above can still be captured as particular cases. The case of local product-market competition corresponds to $\tau = \infty$, which turns equation (13) into equation (8) with $C = N_i$. The case of global product-market competition corresponds to $\tau = 1$, which turns equation (13) into equation (8) with $C = \sum_{j=1}^{I} N_j$. In addition, we can now also consider intermediate cases where $1 < \tau < \infty$.

Regarding interactions, we can think of these as being subject to some spatial decay. Specifically, let us redefine the relevant argument for the interactions function $a(\cdot)$ as the sum of local population and outside population, with the latter adjusted by some decay factor as in Fujita and Ogawa (1982) and Lucas and Rossi-Hansberg (2002). 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 the decay parameter δ measures the strength of across-city relative to within-city interactions ($0 < \delta < 1$). From equation (9), the log productivity of a firm with marginal cost h in city i is given by $\phi = \ln\left[a(N_i + \delta \sum_{j \neq i} N_j)\right] - \ln(h)$. Thus, the gain in log productivity due to interactions in city i (a local measure of the strength of agglomeration) of equation (11) can be redefined as

$$A_i \equiv \ln\left[a(N_i + \delta \sum_{j \neq i} N_j)\right] . \tag{14}$$

The case of local interactions discussed above corresponds to $\delta = 0$, which implies $A_i = \ln[a(N_i)]$. The case of global interactions discussed above corresponds to $\delta = 1$, which implies that $A_i = \ln[a(\sum_{j=1}^{I} N_j)]$. In addition, we can now also consider intermediate cases where $0 < \delta < 1$.

The distribution of firms' log productivities still has its probability density function given by equation (10), which, using subindex i to specify the city, becomes

$$
f_i(\phi) = \begin{cases} 0 & \text{for } \phi < A_i - \ln(\bar{h}_i) \,, \\ \dfrac{e^{A_i - \phi} g\left(e^{A_i - \phi}\right)}{G(\bar{h}_i)} & \text{for } \phi \geqslant A_i - \ln(\bar{h}_i) \,. \end{cases} \tag{15}
$$

In anticipation of the econometric approach developed in the next section, it is also useful write the corresponding cumulative density function, $F_i(\phi)$. To do that compactly, we need to introduce some additional notations. Let

$$
S_i \equiv 1 - G(\bar{h}_i) \tag{16}
$$

denote the proportion of firms that fail to survive product-market competition in city i (a local measure of the strength of selection). To further simplify notation, let us define

$$
\tilde{F}(\phi) \equiv 1 - G\left(e^{-\phi}\right) \tag{17}
$$

as the underlying cumulative density function of log productivities we would observe in all cities in the absence of any selection ($\bar{h}_i = \infty$, $\forall i$) and in the absence of any agglomeration ($A_i = 0$, $\forall i$). Without selection ($\bar{h}_i = \infty$, $\forall i$) all entrants survive regardless of their draw of h. Without agglomeration ($A_i = 0$, $\forall i$), $\phi = -\ln(h)$. Equivalently, $h = e^{-\phi}$. Using the change of variables theorem then yields (17) above. We can then write the cumulative density function of the distribution of log productivities for active firms in city i as

$$
F_i(\phi) = \max\left\{0, \frac{\tilde{F}(\phi - A_i) - S_i}{1 - S_i}\right\} \,. \tag{18}
$$

Relative to the underlying distribution given by (17), agglomeration shifts the distribution rightwards by A_i while selection eliminates a share S_i of entrants (those with lower productivity values). The next section develops an econometric approach to estimate the relative magnitude across cities of agglomeration, as measured by A_i, and selection, as measured by S_i. The following proposition contains our main theoretical result, with predictions for how these expressions vary across cities of different sizes.

Proposition 1. Suppose there are I cities ranked from largest to smallest in terms of population: $N_1 > N_2 > \cdots > N_{I-1} > N_I$, that workers are equally productive in any local firm, that interactions across cities decay by a factor δ, where $0 \leqslant \delta \leqslant 1$, and that selling in a different city raises variable costs by a factor τ, where $1 \leqslant \tau \leqslant \infty$.

  i. Agglomeration leads to the distribution of firms' log productivities being right-shifted by A_i, and if $\delta < 1$ this right shift is greater the larger a city's population: $A_1 > A_2 > \ldots > A_{I-1} > A_I$.

  ii. Firm selection left-truncates a share S_i of the distribution of firms' log productivities, and if $\tau > 1$ this truncation is greater the larger a city's population: $S_1 > S_2 > \ldots > S_{I-1} > S_I$.

iii. If there is no decay in interactions across cities, so that $\delta = 1$, then there are no differences in shift across cities: $A_i = A_j$, $\forall i, j$. 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 Consider any two areas i and j such that $i < j$ (and thus $N_i > N_j$). The extent of the right shift is A_i in city i and A_j in city j. Using equation (14), $0 \leqslant \delta < 1$ directly implies $A_i > A_j$, and $\delta = 1$ implies $A_i = A_j$. 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 (13) for cities i and j can be rewritten:

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

(19)

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

(20)

Subtracting equation (20) from (19) and simplifying yields:

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

(21)

where

$$v(z, \tau) \equiv \int_0^z (z - h)^2 g(h) \mathrm{d}h - \int_0^{z/\tau} (z - \tau h)^2 g(h) \mathrm{d}h \,.$$

(22)

It follows from (21) and $N_i > N_j$ that

$$v(\bar{h}_i, \tau) < v(\bar{h}_j, \tau) \,.$$

(23)

Differentiating (22) with respect to z yields:

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

(24)

If $1 < \tau \leqslant \infty$, then $\partial v(z, \tau)/\partial z > 0$, and thus, by equation (23), $\bar{h}_i < \bar{h}_j$. Hence, by equation (16), $S_i > S_j$. If $\tau = 1$, then by equation (24), $\partial v(z, \tau)/\partial z = 0$, and thus $\bar{h}_i = \bar{h}_j$ and $S_i = S_j$. $\square$

3. Econometric approach

We now develop an econometric approach to estimate the parameters that quantify the importance of selection and agglomeration in the theoretical model for cities of different sizes. The observable information is the cumulative distribution of log productivities in each city. Ideally, we would like to use this information to estimate parameters A_i and S_i from equation (18) for each city. However, this is not possible because the baseline cumulative of log productivities $\tilde{F}$ is not observed. Nevertheless, the following lemma shows that we can get around this issue by comparing the distribution of log productivities across two cities of different sizes i and j to difference out $\tilde{F}$ from equation (18).

Lemma 1. Consider two distributions with cumulative density functions F_i and F_j. Suppose F_i can be obtained by shifting rightwards by A_i some underlying distribution with cumulative density function $\tilde{F}$ and left-truncating a share $S_i \in [0,1)$ of its values:

$$F_i(\phi) = \max\left\{0, \frac{\tilde{F}(\phi - A_i) - S_i}{1 - S_i}\right\} . \tag{25}$$

Suppose F_j can be obtained by shifting rightwards by a different value $A_j \neq A_i$ the same underlying distribution $\tilde{F}$ and left-truncating a different share $S_j \neq S_i$ of its values:

$$F_j(\phi) = \max\left\{0, \frac{\tilde{F}(\phi - A_j) - S_j}{1 - S_j}\right\} . \tag{26}$$

Let

$$A \equiv A_i - A_j , \tag{27}$$

$$S \equiv \frac{S_i - S_j}{1 - S_j} . \tag{28}$$

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

$$F_i(\phi) = \max\left\{0, \frac{F_j(\phi - A) - S}{1 - S}\right\} . \tag{29}$$

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

$$F_j(\phi) = \max\left\{0, \frac{F_i(\phi + A) - \frac{-S}{1-S}}{1 - \frac{-S}{1-S}}\right\} . \tag{30}$$

Proof See the Appendix. □

We are going to use (29) and (30) 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 productivities $\tilde{F}$, which one cannot observe empirically. A limitation is that we are not able to separately identify A_i, A_j, S_i and S_j from the data, but only $A = A_i - A_j$ and $S = (S_i - S_j)/(1 - S_j)$. In other words, we are able to make statements about the relative strength of agglomeration economies in large cities compared to small cities and about the relative strength of firm selection in large cities compared to small cities, but not about the absolute strength of agglomeration economies or firm selection. Parameter A measures how much stronger is the right shift (induced by agglomeration economies in the theoretical model) in city i relative to the smaller city j. In particular, it corresponds to the difference between cities i and j in the strength of agglomeration-induced productivity gains. 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 S measures how much stronger is the left truncation (induced by firm selection in the theoretical model) 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 (29) and (30) 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 uth quantile of F_i and $\lambda_j(u) \equiv F_j^{-1}(u)$ to denote the uth quantile of F_j. If $S > 0$, equation (29) applies and can be rewritten as

$$\lambda_i(u) = \lambda_j\left(S + (1-S)u\right) + A, \quad \text{for } u \in [0,1].\tag{31}$$

If $S < 0$, equation (30) applies and can be rewritten as

$$\lambda_j(u) = \lambda_i\left(\frac{u-S}{1-S}\right) - A, \quad \text{for } u \in [0,1].\tag{32}$$

Making the change of variable $u \to S + (1-S)\,u$ in (32), this becomes

$$\lambda_j\left(S + (1-S)\,u\right) = \lambda_i(u) - A, \quad \text{for } u \in \left[\frac{-S}{1-S}, 1\right].\tag{33}$$

We can then write the following equation that combines (31) and (33):

$$\lambda_i(u) = \lambda_j\left(S + (1-S)u\right) + A, \quad \text{for } u \in \left[\max\left(0, \frac{-S}{1-S}\right), 1\right].\tag{34}$$

Equation (34) cannot be directly used for the estimation because the set of ranks $\left[\max\left(0, \frac{-S}{1-S}\right), 1\right]$ depends on the true value of S, which is not known. We thus make a final change of variable $u \to r_S(u)$, where $r_S(u) = \max\left(0, \frac{-S}{1-S}\right) + \left[1 - \max\left(0, \frac{-S}{1-S}\right)\right]u$, which transforms (34) into

$$\lambda_i\left(r_S(u)\right) = \lambda_j\left(S + (1-S)r_S(u)\right) + A, \quad \text{for } u \in [0,1].\tag{35}$$

Equation (35) 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 agglomeration/shift parameter A and the relative selection/truncation parameter S.

A suitable class of estimators

To estimate A and S, we use the infinite set of equalities given by (35) which can be rewritten in more general terms as $m_\theta(u) = 0$ for $u \in [0,1]$, where $\theta = (A, S)$ and

$$m_\theta(u) = \lambda_i\left(r_S(u)\right) - \lambda_j\left(S + (1-S)r_S(u)\right) - A.\tag{36}$$

We turn to a class of estimators studied by Gobillon and Roux (2008) 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 appendix for details on how these estimators are constructed). We can then introduce an error minimization 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

14

functions, $\langle \cdot, \cdot \rangle$ denote the inner product such that for any functions y and z in $\mathcal{L}^2$, we have: $\langle y,z \rangle = \int_0^1 \int_0^1 y(u)z(v)\mathrm{d}u\mathrm{d}v$, 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: $(B^*By)(v) = \int_0^1 y(u)\ell(v,u)\mathrm{d}u$ and thus $\|By\|^2 = \int_0^1 \int_0^1 y(u)\ell(v,u)y(v)\mathrm{d}u\mathrm{d}v$. Let $n = (n_i,n_j)'$, where n_i and n_j denote respectively the number of observations of the distributions F_i and F_j. The vector of parameters θ can then be estimated as

$$\hat{\theta} = \arg\min_\theta \|B_n \hat{m}_\theta\| \,, \tag{37}$$

where B_n is a sequence of bounded linear operators.[11] Gobillon and Roux (2008) show that if $\hat{\lambda}_i$ and $\hat{\lambda}_j$ are some appropriate estimators that are consistent, asymptotically normal and continuous twice-differentiable, and if the sequence B_n is such that $\|B_n \hat{m}_\theta\| \to \|B m_\theta\|$ as $\min(n_i,n_j) \to +\infty$, where B is a bounded linear operator verifying $B m_\theta = 0 \implies m_\theta = 0$, then the estimator $\hat{\theta}$ defined by (37) is consistent and asymptotically normal. They also show how to determine the weights $\ell(v,u)$ leading to the optimal estimator.

Implementation

While it is possible to compute the weights leading to the optimal estimator, they cannot be used in practice because they depend on the true value of the parameters θ. Alternatively, we could rely on a simple weighting 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 would simplify to:

$$\hat{\theta} = \arg\min_\theta \left(\int_0^1 [\hat{m}_\theta(u)]^2 \mathrm{d}u \right) \,. \tag{38}$$

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 (35) for this reverse comparison. Making the change of variable $u \to \frac{u-S}{1-S}$ in (31), this becomes

$$\lambda_j(u) = \lambda_i \left(\frac{u-S}{1-S} \right) - A \,, \quad \text{for } u \in [S,1] \,. \tag{39}$$

We can then write the following alternative equation to (34) that combines (32) and (39):

$$\lambda_j(u) = \lambda_i \left(\frac{u-S}{1-S} \right) - A \,, \quad \text{for } u \in [\max(0,S),1] \,. \tag{40}$$

---

[11]The 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) $u_i,\ldots,u_K$ such that the system $m_\theta(u_i) = 0$ for $i = 1,\ldots,K$ admits a unique solution in θ.

Let $\tilde{r}_S(u) = \max(0, S) + [1 - \max(0, S)]\, u$. With a final change of variable $u \to \tilde{r}_S(u)$ on (40), this provides a new set of equalities $\tilde{m}_\theta(u) = 0$, for $u \in [0, 1]$, where

$$\tilde{m}_\theta(u) = \lambda_j\left(\tilde{r}_S(u)\right) - \lambda_i\left(\frac{\tilde{r}_S(u) - S}{1 - S}\right) + A\ . \tag{41}$$

Let $\hat{\tilde{m}}_\theta(u)$ 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 $\hat{\lambda}_j$ (see the appendix for details). The estimator we actually use is then

$$\hat{\theta} = \arg\min_\theta M(\theta)\ , \quad \text{where} \quad M(\theta) = \int_0^1 [\hat{m}_\theta(u)]^2 \mathrm{d}u + \int_0^1 [\hat{\tilde{m}}_\theta(u)]^2 \mathrm{d}u\ . \tag{42}$$

In the results below, we report $\hat{\theta} = (\hat{A}, \hat{S})$, as well as a measure of goodness of fit $R^2 = 1 - \frac{M(\hat{A}, \hat{S})}{M(0, 0)}$. Standard errors of the estimated parameters are bootstrapped drawing some observations 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 estimations

Data

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 productive and financial assets to compute a measure of capital. This is done using the sum of 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 the firm to which the establishment belongs (for matching with BRN-RSI), and the municipality where the establishment is located. We allocate each municipality to its metropolitan area ('Aire Urbaine') when it is part of one.

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.

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.[12] We retain information on all establishments from all firms with 6 employees or more in all manufacturing sectors and in business services, with the noted exception of finance and insurance. We end up with data on 153,130 firms and 203,300 establishments observed at least once during the study period.

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 skilled levels, and use $l_{s,t}$ to denote the share of the firm's workers with skilled level s:

$$V_t = (k_t)^{\beta_1} \left(l_t \sum_{s=1}^{3} \varsigma_s l_{s,t} \right)^{\beta_2} e^{\beta_{3,t} + \phi_t} , \tag{43}$$

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

$$\ln(V_t) = \beta_1 \ln(k_t) + \beta_2 \ln(l_t) + \beta_2 \ln \left(\sum_{s=1}^{3} \varsigma_s l_{s,t} \right) + \beta_{3,t} + \phi_t . \tag{44}$$

To linearise (44), 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 $l_{s,t} \approx \zeta_s$, then

$$\beta_2 \ln \left(\sum_{s=1}^{3} \varsigma_s l_{s,t} \right) \approx \beta_2 \left[\ln \left(\sum_{s=1}^{3} \varsigma_s \zeta_s \right) - 1 \right] + \sum_{s=1}^{3} \sigma_s l_{s,t} , \tag{45}$$

where $\sigma_s \equiv \beta_2 \varsigma_s / (\sum_{s=1}^{3} \varsigma_s \zeta_s)$. Substituting equation (45) into (44) yields:

$$\ln(V_t) = \beta_{0,t} + \beta_1 \ln(k_t) + \beta_2 \ln(l_t) + \sum_{s=1}^{3} \sigma_s l_{s,t} + \phi_t , \tag{46}$$

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

We obtain log TFP by estimating equation (46) separately for each sector in level 2 of the Nomenclature Economique de Synthèse (NES) sectoral classification, which leaves us with 16

---

[12]The merged data set contains much more information than is usually available. For instance, US-based research relies either on sectoral surveys or on five-yearly censuses for which value added is difficult to compute. We instead have exhaustive annual data. We also have information on the number of hours worked by skill level instead of total employment as is often the case.

manufacturing sectors and business services.[13] 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,

$$\hat{\phi} = \frac{1}{T} \sum_{t=1}^{T} \hat{\phi}_t \, , \tag{47}$$

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

For our baseline results, we estimate equation (46) 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 the Appendix.

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 85 percent of firms, 68 percent of establishments, and 46 percent of employment). Later, we take advantage of establishment-level data on hours worked by skill and 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 cities where it has establishments, separately for each sector:

$$\hat{\phi} = \sum_{i=1}^{I} \nu_i l_i + \epsilon \, , \tag{48}$$

where i indices metropolitan areas (there are 364 in France), and l_i denotes the share of a firm's labour (in hours worked) in area i, averaged over the period 1994–2002. Parameter ν_i is common to all firms and establishments in area 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. Baseline results

The quantile approach described in Section 3 estimates the amount of left truncation and right shift that, when applied to one distribution of firms' log productivities, best approximate another distribution of firms' log productivities. To implement the approach, after estimating TFP for mono-establishment firms as described in Section 4 using OLS, we must choose which two distributions to compare. For our baseline estimates, we lump urban areas together based on their population size. In particular, we compare the distribution of firms' log productivities in urban

---

[13]Unfortunately, we cannot include banking and insurance in our estimation because the location of establishments is not available for these sectors, which have distinct reporting rules. We also exclude distribution and consumer services from our main estimations. The assignment of a specific location to distribution (which involves moving goods across locations) is difficult and the estimation of a production function in consumer services is more problematic (but see the bottom-right panel of Figure 2 for an illustration from consumer services).

Table 1: Baseline estimation results, cities with pop.> 200,000 vs. pop.< 200,000

| Sector | OLS, mono-establishments | | | | | | | |
| | A | S | R^2 | A | R^2 | S | R^2 | obs. |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Food, beverages, tobacco	0.06 (0.01) *	0.01 (0.00) *	0.90	0.07 (0.01) *	0.82	0.03 (0.01) *	0.40	22,049
Apparel, leather	0.09 (0.04) *	-0.04 (0.06) *	0.42	0.05 (0.01) *	0.18	-0.01 (0.00) *	0.06	5,804
Publishing, printing, recorded media	0.20 (0.01) *	-0.03 (0.01) *	0.85	0.17 (0.01) *	0.73	0.00 (0.02)	0.00	9,236
Pharmaceuticals, perfumes, soap	0.09 (0.05)	-0.04 (0.04)	0.76	0.02 (0.04)	0.01	-0.02 (0.01)	0.58	1,069
Domestic appliances, furniture	0.14 (0.01) *	0.00 (0.01)	0.82	0.14 (0.01) *	0.81	0.03 (0.02) *	0.09	6,362
Motor vehicles	0.10 (0.02) *	-0.01 (0.02)	0.54	0.08 (0.02) *	0.41	0.00 (0.01)	0.02	1,442
Ships, aircraft, railroad equipment	0.11 (0.03) *	-0.02 (0.01)	0.75	0.08 (0.03) *	0.45	-0.01 (0.02)	0.08	1,016
Machinery	0.08 (0.01) *	-0.01 (0.00) *	0.97	0.08 (0.01) *	0.93	0.01 (0.01)	0.03	14,736
Electric and electronic equipment	0.08 (0.01) *	-0.01 (0.00)	0.97	0.08 (0.01) *	0.94	0.02 (0.02)	0.03	5,749
Building materials, glass products	0.07 (0.02) *	0.00 (0.01)	0.83	0.06 (0.01) *	0.81	0.00 (0.01)	0.04	3,196
Textiles	0.06 (0.02) *	0.00 (0.01)	0.61	0.06 (0.02) *	0.57	0.00 (0.01)	0.00	3,365
Wood, paper	0.10 (0.01) *	-0.01 (0.01)	0.91	0.09 (0.01) *	0.85	0.01 (0.01)	0.02	5,872
Chemicals, rubber, plastics	0.09 (0.01) *	0.00 (0.01)	0.95	0.09 (0.01) *	0.95	0.01 (0.01)	0.14	5,337
Basic metals, metal products	0.08 (0.01) *	0.00 (0.00)	0.97	0.08 (0.01) *	0.96	0.01 (0.01) *	0.07	14,305
Electric and electronic components	0.08 (0.02) *	0.00 (0.01)	0.94	0.08 (0.02) *	0.93	0.01 (0.01)	0.15	2,579
Consultancy, advertising, business services	0.19 (0.01) *	-0.02 (0.01) *	0.95	0.17 (0.01) *	0.88	0.09 (0.03) *	0.06	37,041
All sectors	0.11 (0.00) *	-0.02 (0.00) *	0.71	0.09 (0.00) *	0.57	0.00 (0.00) *	0.00	139,143

areas with over 200,000 people with the corresponding distribution in urban areas with less than 200,000 people and rural areas. Later, we report as robustness checks results comparing urban areas using alternative population size classes, results comparing individual urban areas, results using employment areas instead or urban areas, split by employment density instead of population size, and results using alternative methods to estimate TFP.[14]

Columns (1) and (2) in Table 1 show our baselines estimates of A and S for two-digit manufacturing and business service sectors. Recall that $S \equiv \frac{S_i - S_j}{1 - S_j}$, where i corresponds to large cities and j corresponds to small cities. If $S > 0$, then the strength of selection increases with city size. If $S = 0$, then the strength of selection does not vary with city size. Recall also that $A \equiv A_i - A_j$. If $A > 0$, then the strength of agglomeration increases with city size. If $A = 0$, then the strength of agglomeration does not vary with city size. Column (3) in Table 1 reports a pseudo-R^2 as defined in Section 3.

In column (1), A is always positive. Statistical significance at the 5 percent level is marked with an asterisk next to the bootstrapped standard errors reported in parenthesis. A is significantly different from zero in all cases but one. For all sectors it takes a value $A = 0.11$, which implies a 12 percent productivity increase. These results suggest that agglomeration economies are stronger in large cities than in small cities. Our model shows that the extent to which agglomeration economies vary across cities of different size 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 empirical literature looking at the spatial decay of different types of agglomeration economies (see Rosenthal and Strange, 2004). We postpone the discussion of the economic significance of our findings until later.

For 11 sectors out of 16, S is not statistically different from zero. It is negative and significant in four sectors and for all sectors pooled together.[15] It is positive and significant in one sector only. In all cases, however, S remains small. This suggests that there is not much difference between large and small cities in the strength of selection. Note that this does not imply that selection is not important. It simply suggests that its importance is similar in cities of different sizes. Our model shows that the extent to which selection varies across cities of different sizes is closely related to the extent to which product market competition is local or global (national in this case). Our results are consistent with a situation where French firms compete with similar intensity on national markets regardless of their location.

Column (4) in Table 1 reports our estimates of A when we impose the restriction $S = 0$ (no difference in the strength of selection between large and small cities). Not surprisingly given how close to zero the estimates of S in column (2) are, the estimates of A in column (4) are very close

---

[14]Whenever 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 truncation and shift, 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. While our estimates lose some precision when we trim fewer outliers, results are qualitatively unchanged.

[15]Having negative estimates of S, even if small in absolute value, may seem strange in light of the model. We argue in Section 7 that it is an artifact of the simple way in which we capture agglomeration economies in these baseline results. As we shall see, once we generalise our approach, the estimates of S become zero.

PDF

-1.0 -0.5 0.0 0.5 1.0 log TFP

Panel (a)
Electric generators, engines and transformers
(truncation and no shift
at large vs. small cities)

Panel (b)
Fabricated metal products
(no truncation and shift
at large vs. small cities)

Panel (c)
Paints, agro-chemicals, other chemicals
(truncation and shift
at large vs. small cities)

Panel (d)
Hairdressing, beauty, funeral services
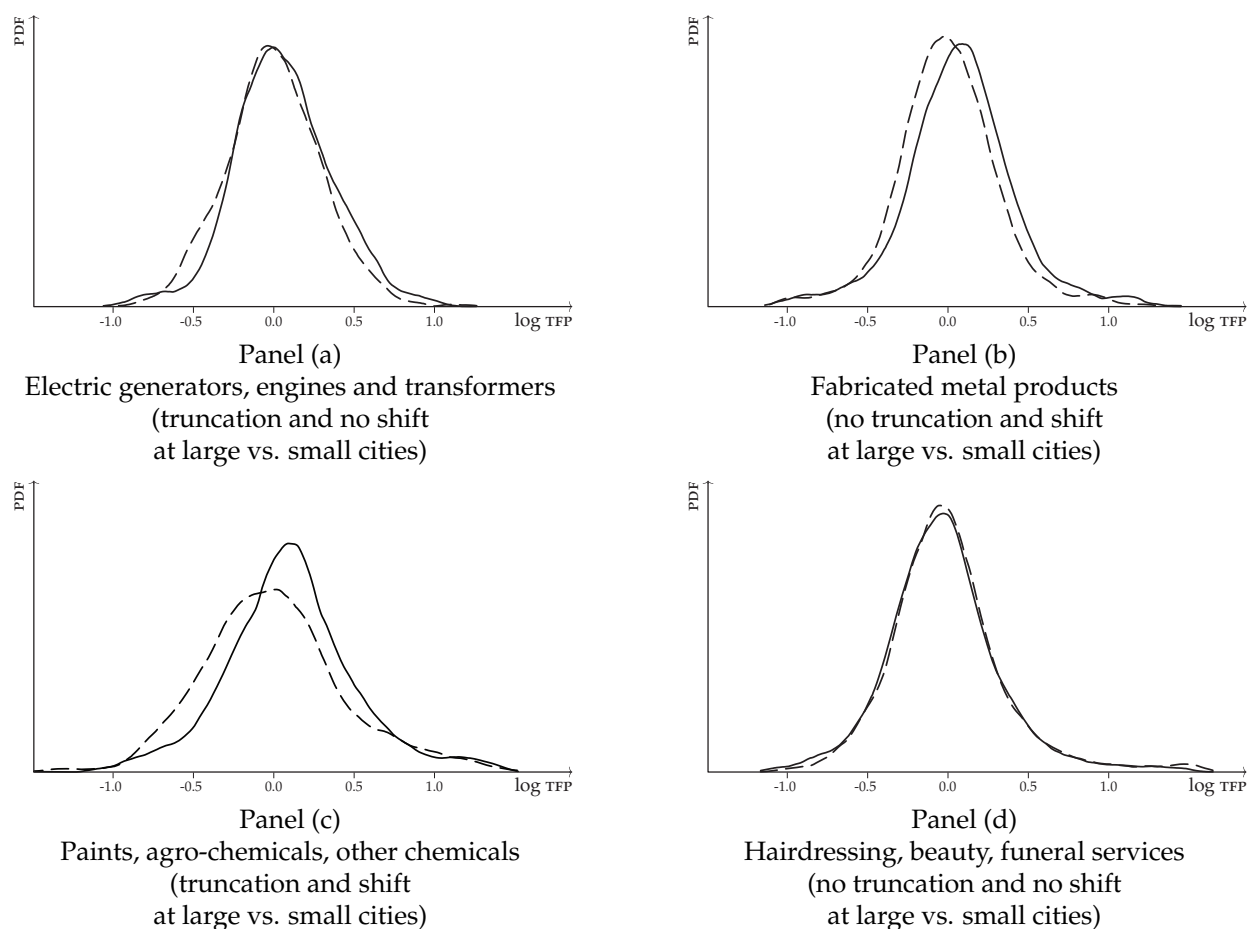(no truncation and no shift
at large vs. small cities)

Figure 2: Empirical log productivity distributions in large (solid) and small cities (dashed)

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).

Column (6) in Table 1 reports our estimates of S when we impose the restriction $A = 0$ (no difference in the strength of agglomeration between large and small cities). 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 agglomeration to vary across cities of different sizes, we pick up part of the agglomeration effects as variation in selection. Column (7) reports the pseudo-R^2 under the restriction $A = 0$. A comparison with column (3) shows that fit deteriorates very 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 agglomeration economies are stronger in large cities than in small cities and the absence of significant differences in selection effects.[16]

---

[16]Other estimation methods for TFP, such as the procedures developed by Olley and Pakes (1996) and Levinsohn and Petrin (2003) and cost shares lead to similar conclusions. Considering all establishments as opposed to only mono-establishment firms also yields very similar results. Finally, our results are not affected by trimming 2 percent instead of 1 percent of observations at both extremes of the log productivity distribution. We do not report these results for our baseline estimation. They are available upon request. Below we report results using these alternative estimation methods and alternative samples of establishments to assess the robustness of our main results.

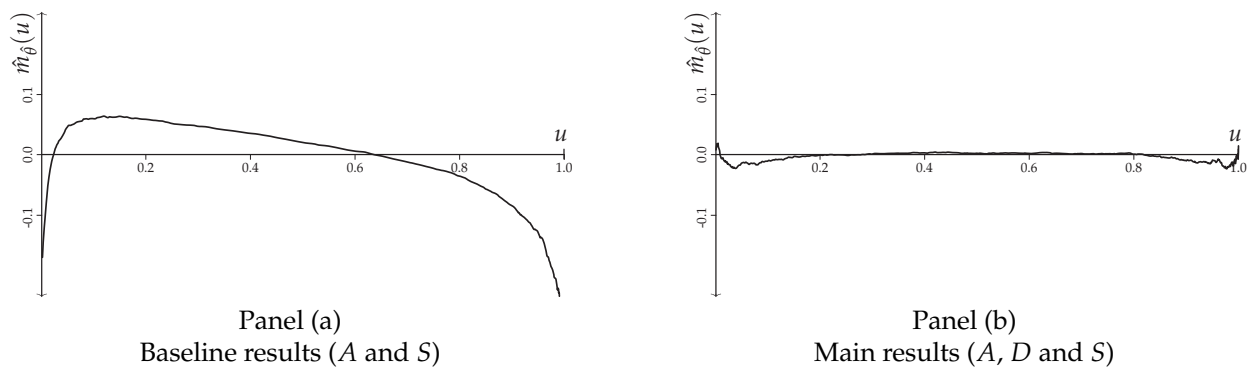| Panel (a)	Panel (b)
Baseline results (A and S)	Main results (A, D and S)

Figure 3: Estimation errors by quantile

We also find broadly similar results for more finely defined sectors, i.e., sectors as defined by the level 3 of the NES classification. There are two differences however. First, because these more disaggregated sectors are sometimes small, differences between small and large cities are less frequently statistically significant than with two-digit sectors. Second, although the general trend in the data is stronger agglomeration in larger cities and no difference in selection, looking at more finely defined sectors shows clearly that this pattern is not universal. Figure 2 represents the distributions of log productivities in small and large cities for four three-digit sectors. These closely match the four theoretical patterns of Figure 1. Electric generators, engines and transformers (Panel a) is a rare example of a sector exhibiting stronger selection effects in large cities but no difference in agglomeration effects. Fabricated metal products (Panel b) illustrates the opposite and more common pattern of no difference in selection and stronger agglomeration in large cities. Paints, agro-chemicals, and other chemicals (Panel c) is a case of both stronger agglomeration and stronger selection effects in large cities. Finally, hairdressing, beauty, and funeral services (Panel d) provides, as might have been expected, an example of a sector where the distribution of firms' log productivities is almost exactly the same in small and large cities.

While the pseudo-R^2 of column (3) is high (including above 0.90 in a majority of sectors), looking at the fit of the estimation in more detail suggests that our baseline analysis does not fully explain the differences between the distributions of log productivities in large and small cities. Panel (a) in Figure 3 (ignore Panel b for now) provides some insight into what is missing from our model and empirical approach as we have presented it so far. The graph plots, for all sectors combined (bottom row of Table 1), the values of $\hat{m}_{\hat{\theta}}(u)$. That is, the Figure plots for each quantile (given by a point on the horizontal axis) the difference between its value in the large city distribution and the value that results from shifting and truncating the small-city log-productivity distribution using the estimated values of A and S. There is very marked pattern, where errors tend to be positive for the lower quantiles and negative for the higher quantiles. This indicates that, by forcing all establishments to have the same productivity boost from locating in a large city, we are giving establishments at the lower end of the productivity distribution too large a boost and establishments at the upper end of the productivity distribution too small a boost. In other words, the figure indicates that, contrary to what we have assumed so far, more productive establishments

benefit more from agglomeration. We now extend our model and empirical approach to allow for this.[17]

6. When more productive establishments benefit more from agglomeration

So far we have taken the simple view that agglomeration economies raise the log productivity of all establishments in larger cities by the same amount. We now generalise our theoretical and empirical frameworks to allow the magnitude of agglomeration economies to be systematically related not just to city size but also to individual productivity. In particular, we conjecture that, while agglomeration economies raise the productivity of all firms in larger cities, they raise the productivity of the most productive firms to a greater extent.

Extending the model

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 $a(N_i + \delta \sum_{k \neq i} N_k) h^{-(D_i - 1)}$, where

$$D_i \equiv \ln \left[d(N_i + \delta \sum_{k \neq i} N_k) \right] , \tag{49}$$

$d(0) = 1$, $d' > 0$ 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 given by

$$\phi = \ln \left(\frac{Q}{l} \right) = A_i - D_i \ln(h) . \tag{50}$$

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

$$F_i(\phi) = \max \left\{ 0, \frac{\tilde{F} \left(\frac{\phi - A_i}{D_i} \right) - S_i}{1 - S_i} \right\} . \tag{51}$$

Relative to the underlying log productivity distribution $\tilde{F}$, agglomeration both dilates the distribution by a factor D_i and shifts the distribution rightwards by A_i, while selection eliminates a share S_i of entrants (those with lower productivity values). Proposition 1 can then be rewritten as follows.

Proposition 1′. Suppose there are I cities ranked from largest to smallest in terms of population: $N_1 > N_2 > \cdots > N_{I-1} > N_I$, that workers are more productive when they work for a more

---

[17]It is also worth noting that, regardless of the general downward-sloping error pattern, errors for the first few quantiles are negative before quickly becoming positive above the first 2 percent of quantiles. This is a sign that the small negative value estimated for S ($S = -0.02$) leads to a very bad fit at the very bottom of the distribution even if it helps improve the overall fit. We return to this issue below, where we argue that not allowing more productive establishments to get an additional productivity boost in large cities results in a downward bias for S as well as in an upward bias for A.
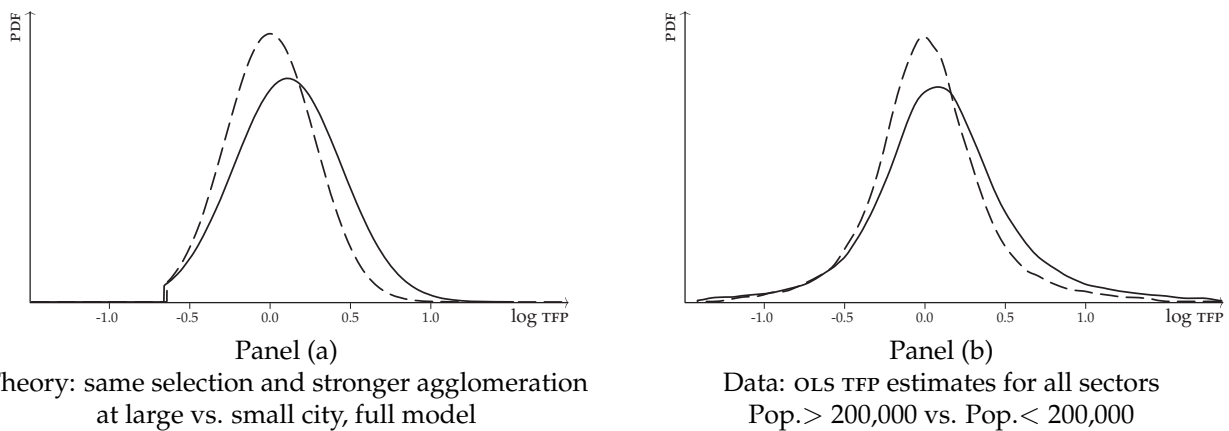
Figure 4: Log-productivity distributions in large (solid) and small cities (dashed)

efficient (lower h) firm and that this effect is enhanced by interactions, that interactions across cities decay by a factor δ, where $0 \leqslant \delta < 1$, and that selling in a different city raises variable costs by a factor τ, where $1 \leqslant \tau \leqslant \infty$.

i. Agglomeration leads to the distribution of log productivities 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: $D_1 > D_2 > \ldots > D_{I-1} > D_I$ and $A_1 > A_2 > \ldots > A_{I-1} > A_I$.

ii. Firm selection left-truncates a share S_i of the distribution of log productivities, and if $\tau > 1$ this truncation is greater the larger a city's population: $S_1 > S_2 > \ldots > S_{I-1} > S_I$.

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 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 $0 \leqslant \delta < 1$, by equation (49), $D_i > D_j$ and, by equation (14), $A_i > A_j$. If instead $\delta = 1$, by the same two equations, $D_i = D_j$ and $A_i = A_j$. The proportion of truncated values of $\tilde{F}$ is S_i in city i and S_j in city j. The free entry conditions of equations (19) and (20) still apply. If $1 < \tau \leqslant \infty$, by equation (24), $\bar{h}_i < \bar{h}_j$ and thus, by equation (16), $S_i > S_j$. If $\tau = 1$, by the same two equations, $\bar{h}_i = \bar{h}_j$ and thus $S_i = S_j$. $\qquad\square$

Panel (a) of Figure 4 re-draws Panel (b) of Figure 1 for the full model. It plots the distribution of firms' log productivities in a city with a large population (solid line) and in a city with a small population (dashed line) with global product-market competition and local interactions (i.e., when there are stronger agglomeration economies in large cities than in small cities but the same strength of selection in both). The difference with respect to Panel (b) of Figure 1 is that more productive firms now benefit even more from locating in large cities, so that stronger agglomeration economies get reflected in both a right shift and a dilation of the log-productivity distribution. This can be

24

seen graphically in the peak for the large city being lower and in the gap between the distributions getting larger as we move towards the right. Panel (b) of Figure 4 plots the actual distribution of log productivities in large cities (urban areas with over 200,000 people, solid line) and small cities (urban areas with less than 200,000 people and rural areas, dashed line) for all sectors together using OLS TFP estimates. We can see it looks remarkably similar to the theoretical benchmark in Panel (a) of Figure 4. To show more formally that the extended model is better at capturing differences in productivity between cities of different sizes, we now extend our econometric approach.

Extending the econometric approach

We now show how to incorporate dilation into our econometric approach. The following generalization of Lemma 1 shows that, by comparing once again the distribution of log productivities across two cities of different sizes i and j, we can difference out $\tilde{F}$ from equation (51).

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:

$$F_i(\phi) = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi-A_i}{D_i}\right) - S_i}{1 - S_i}\right\}. \tag{52}$$

Suppose F_j can be obtained by dilating by a different factor $D_j \neq D_i$ and shifting rightwards by a different value $A_j \neq A_i$ the same underlying distribution $\tilde{F}$ and also left-truncating a different share $S_j \neq S_i$ of its values:

$$F_j(\phi) = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi-A_j}{D_j}\right) - S_j}{1 - S_j}\right\}. \tag{53}$$

Let

$$D \equiv \frac{D_i}{D_j}, \tag{54}$$

$$A \equiv A_i - DA_j, \tag{55}$$

$$S \equiv \frac{S_i - S_j}{1 - S_j}. \tag{56}$$

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:

$$F_i(\phi) = \max\left\{0, \frac{F_j\left(\frac{\phi-A}{D}\right) - S}{1 - S}\right\}. \tag{57}$$

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

$$F_j(\phi) = \max\left\{0, \frac{F_i\left(D\phi + A\right) - \frac{-S}{1-S}}{1 - \frac{-S}{1-S}}\right\}. \tag{58}$$

Proof See the Appendix. □

To estimate the set of parameters $\theta = (A, D, S)$, we first rewrite (57) and (58) in quantiles. If $S > 0$, equation (57) applies and can be rewritten as

$$\lambda_i(u) = D\lambda_j \left(S + (1 - S)u \right) + A , \quad \text{for } u \in [0, 1] . \tag{59}$$

If $S < 0$, equation (58) applies and can be rewritten as

$$\lambda_j(u) = \frac{1}{D}\lambda_i \left(\frac{u - S}{1 - S} \right) - \frac{A}{D} , \quad \text{for } u \in [0, 1] . \tag{60}$$

Performing on (57) and (58) exactly the same steps we performed on (29) and (30) to obtain (36) and (41) yields:

$$m_\theta(u) = \lambda_i \left(r_S(u) \right) - D\lambda_j \left(S + (1 - S)r_S(u) \right) - A , \tag{61}$$

$$\tilde{m}_\theta(u) = \lambda_j \left(\tilde{r}_S(u) \right) - \frac{1}{D}\lambda_i \left(\frac{\tilde{r}_S(u) - S}{1 - S} \right) + \frac{A}{D} . \tag{62}$$

The estimator we use is still given by (42), where $\hat{m}_\theta(u)$ and $\hat{\tilde{m}}_\theta(u)$ still denote the empirical counterparts of $m_\theta(u)$ and $\tilde{m}_\theta(u)$, now redefined in (61) and (62).

7. Main empirical results

For our main results, we rely again on TFP estimated as described in Section 4. Relative to our baseline results, we now estimate how the distribution of firms' log productivities in large cities is best approximated by shifting, dilating and truncating the distribution of firms' log productivities in small cities. We now estimate a shift parameter, A, a dilation parameter, D, and a truncation parameter, S, comparing large cities (urban areas with over 200,000 people) and small cities (urban areas with less than 200,000 people and rural areas) for 16 manufacturing and business service sectors and all sectors together.

Columns (1), (2), and (3) of Table 2 present our main results for A, D, and S. Recall that, in the extended model that serves as a basis for these main results, greater agglomeration economies in large cities result in the distribution of log productivities in large cities being both right shifted and dilated relative to the distribution in small cities, i.e., in $A > 0$ and $D > 1$. The value of A corresponds to the average increase in log productivity that would arise in large cities relative to small cities absent any selection.[18] When $A > 0$, values of D above unity are evidence that agglomeration economies in large cities benefit more the more productive firms, whereas values of D below unity would indicate that agglomeration economies benefit less the more productive firms. As in the baseline, positive values of S correspond to the distribution of firms' log productivities in large cities being more truncated than in small cities, whereas negative values correspond to more truncation in small cities.

---

[18]Note that we normalise our log-TFP estimates so that our estimates of A in Table 2 are directly comparable with those of Table 1. This involves choosing units of value added so that average log-TFP in small cities is zero, which affects neither D nor S.

Table 2: Main estimation results, cities with pop.$> 200{,}000$ vs. pop.$< 200{,}000$

Sector	OLS, mono-establishments				
	A	D	S	R^2	obs.
	(1)	(2)	(3)	(4)	(5)
Food, beverages, tobacco	0.07 (0.00) *	0.94 (0.02) *	0.00 (0.00)	0.95	22,049
Apparel, leather	0.04 (0.01) *	1.37 (0.05) *	0.01 (0.01)	0.98	5,804
Publishing, printing, recorded media	0.17 (0.01) *	1.26 (0.04) *	0.00 (0.00)	0.98	9,236
Pharmaceuticals, perfumes, soap	0.05 (0.05)	1.19 (0.11)	-0.01 (0.04)	0.91	1,069
Domestic appliances, furniture	0.13 (0.01) *	1.22 (0.04) *	0.01 (0.01) *	0.98	6,362
Motor vehicles	0.08 (0.03) *	1.29 (0.14) *	0.01 (0.03)	0.80	1,442
Ships, aircraft, railroad equipment	0.10 (0.03) *	1.09 (0.18)	-0.01 (0.03)	0.81	1,016
Machinery	0.08 (0.01) *	1.04 (0.03)	0.00 (0.00)	0.98	14,736
Electric and electronic equipment	0.08 (0.01) *	1.02 (0.05)	0.00 (0.01)	0.97	5,749
Building materials, glass products	0.06 (0.02) *	1.06 (0.06)	0.00 (0.02)	0.89	3,196
Textiles	0.05 (0.01) *	1.14 (0.08) *	0.00 (0.01)	0.92	3,365
Wood, paper	0.09 (0.01) *	1.11 (0.05) *	0.00 (0.01)	0.99	5,872
Chemicals, rubber, plastics	0.08 (0.01) *	1.04 (0.04)	0.00 (0.01)	0.96	5,337
Basic metals, metal products	0.08 (0.01) *	1.06 (0.02) *	0.00 (0.00)	1.00	14,305
Electric and electronic components	0.08 (0.02) *	0.99 (0.07)	0.00 (0.03)	0.94	2,579
Consultancy, advertising, business services	0.19 (0.02) *	1.05 (0.04)	-0.01 (0.03) *	0.96	37,041
All sectors	0.09 (0.00) *	1.22 (0.01) *	0.00 (0.00)	1.00	139,143

In column (1) of Table 2, A is always positive and, like in Table 1, it is significant at 5 percent in all cases but one. The only difference with Table 1 is that the estimates for A are now slightly lower. For instance, when considering all sectors, we find $A = 0.09$ in Table 2 whereas $A = 0.11$ for the baseline in Table 1. Column (2) in Table 2 reports our estimates of D. In eight sectors D is statistically different from unity. For seven of these sectors (and for all sectors together), D is above unity. In only one of these sectors, D is below unity. In the other sectors, D is not statistically different from unity although the point estimates are usually above one. There is thus a tendency for the distribution of firms' log productivities to be more dilated in larger cities for about half the sectors and for all sectors combined. With $A > 0$, the finding that D is often above unity is indicative that it is the most productive firms that benefit the most from agglomeration. For all sectors, $A = 0.09$ and $D = 1.22$ imply that firms are on average 9 percent more productive in large cities but that this productivity advantage is 14 percent for firms at the first quartile and only 5 percent for firms at the bottom quartile.[19]

Turning to S in column (2), there is only one case of a sector, domestic appliances and furniture, with a positive and significant value for S, although this value is small at 0.01. There is also only one case, consultancy, advertising, and business services, with a negative and significant value for S, again small at -0.01. In all other cases, the estimated value of S is not significantly different from zero. This lack of significance is not due to imprecise estimates. On the contrary, the standard errors for S are small, like the standard errors for A. Adding to this, we note that in 11 cases out of 17 (including all sectors combined), the estimated value for S is precisely 0.00. These results provide even stronger evidence than our baseline results that there are no differences between small and large cities in the truncation of the distribution of firms' log productivities. Market selection appears to have a similar intensity across cities in France irrespective of their size.

To summarise, firms are more productive in large cities. 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 agglomeration economies boost the productivity of all firms, and in about half of the sectors this increase in productivity is strongest for the most productive firms.

More generally, a comparison between Tables 1 and 2 suggests that it is important to allow for more productive firms to benefit more from agglomeration. When one fails to do so, as in our baseline results of Table 1, estimates 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 1 and 2 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 0.42 to 0.98 when adding D to the estimation. Overall, the fit in column (4) of Table 2 is very good. The pseudo-R^2 is always above

---

[19]By shifting, dilating and truncating the small city distribution using the estimated values of A, D, and S, we obtain a predicted productivity advantage of 4.6 percent for firms at the bottom quartile and 14.1 percent for firms at the top quartile. In the empirical distribution for large cities, we find that these advantages are 4.7 percent and 13.9 percent, respectively. For the bottom and top deciles of the large city distribution and relative to the small city distribution, we find productivity advantages of 0.9 percent and 21.1 percent, respectively.

Table 3: Robustness, cities with pop.$>$ 200,000 vs. pop.$<$ 200,000, alternative estimation methods and alternative samples

Method	A	D	S	R^2	obs.
	(1)	(2)	(3)	(4)	(5)
	all sectors, mono-establishments				
Ordinary Least Squares	0.09 (0.00) *	1.22 (0.01) *	0.00 (0.00)	1.00	139,143
Olley-Pakes	0.08 (0.01) *	1.10 (0.04) *	0.00 (0.00) *	0.98	56,920
Levinsohn-Petrin	0.09 (0.00) *	1.10 (0.01) *	0.00 (0.00)	1.00	101,714
Cost shares	0.07 (0.00) *	1.20 (0.01) *	0.00 (0.00)	0.98	139,143
	all sectors, all establishments				
Ordinary Least Squares	0.09 (0.00) *	1.12 (0.00) *	0.00 (0.00) *	1.00	199,235
Olley-Pakes	0.10 (0.00) *	1.10 (0.00) *	0.01 (0.00) *	0.99	97,246
Levinsohn-Petrin	0.11 (0.00) *	1.04 (0.00) *	0.00 (0.00) *	0.93	155,106
Cost shares	0.08 (0.00) *	1.10 (0.00) *	0.00 (0.00) *	1.00	199,235

0.80 and it is even above 0.95 in a majority of sectors and very close to 1.00 for all sectors combined.

The importance of accounting for the greater benefit of agglomeration economies for large firms is perhaps most clearly seen when one compares the plots of estimation errors by quantile in the two panels of Figure 3. Panel (a) corresponds to our baseline results, and it was the clear downward-sloping pattern in this panel that lead us to add parameter D to the estimation. Panel (b) corresponds to our main results and plots the difference between the value of each quantile in the large city distribution and the value of the same quantile in the distribution that results from shifting, dilating, and truncating the small-city log-productivity distribution using the estimated values of A, D, and S (the estimated value of S being in fact zero, thus leading to no truncation). Estimation errors are greatly reduced relative to those of Panel (a) and the clear downward-sloping pattern apparent in Panel (a) is gone. In fact errors in Panel (b) are almost uniformly zero except for a little wiggle at the both extremes, where productivity values are more scattered and the fit between the distributions inevitably loses precision.

Robustness

To assess the strength of the findings of Table 2, we now turn to a series of robustness checks. First, one might question our TFP estimates. 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. In the top half of Table 3, we report results for all sectors combined using four alternative methods to estimate TFP. The first line of results in Table 3 reports the same OLS results as in the last line of Table 2 to ease comparisons. The next line 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 is very similar to its corresponding OLS value, 0.08 instead of 0.09. The estimates of S are also very close tot 0.00 in both cases. The only difference when using Olley-Pakes is that the tiny amount of truncation (the estimated parameter is $S = 0.004$) is statistically significant whereas it is insignificant with OLS. Finally, there is a small difference for the dilation parameter D. It is equal to 1.10 when using Olley-Pakes against 1.22 with OLS. Overall the differences between OLS and Olley-Pakes TFP are small and could be due to the substantially smaller sample size with Olley-Pakes. The number of establishments used for the estimation drops from 139,143 with OLS to 56,920 with Olley-Pakes. This is due to the need to observe establishments over time in the latter to compute investment. Estimating TFP using the method proposed by Levinsohn and Petrin (2003) in the third line of results in Table 3 yields estimates that are very similar to those of Olley-Pakes TFP. The fourth line reproduces the estimation of A, D, and S when the underlying TFP is estimated using a simple cost-share approach.[20] The results are again very similar. 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 2.

The bottom half of Table 3 replicates the same four estimations of A, D and S as the first panel but this time considering all establishments, affiliated and unaffiliated, instead of only mono-establishment firms. For OLS TFP, the results are very close to those with mono-establishment firms, except for less dilation in big cities. The next three lines report results for the alternative approaches to TFP estimations as described above. The results are very close to their corresponding results for mono-establishments firms in the first panel of the same table. They are also close to those obtained with OLS TFP. Overall we conclude that neither the sample of establishment we use nor the specific method we implement to estimate TFP have much bearing on our results.

Table 4 reports additional results using first a different zoning and then different urban groupings. The first line reproduces again our main results comparing urban areas with over 200,000 people and urban areas with less than 200,000 people and rural areas for all sectors together. The second line of results in that table repeats the same estimation of A, D, and S (OLS, all sectors, mono-establishments), but compares employment areas with above and below median employment density instead of urban areas with above and below 200,000 people. Employment area boundaries are drawn to capture local labour markets on the basis of commuting patterns whereas urban area boundaries are drawn to capture cities. 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 very similar. The fit is also very good. Table 5 in Appendix D reports

---

[20]We 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 very localised markets. It is not suitable for a broad cross-section of sectors nor when pulling all sectors together.

Table 4: Robustness, alternative comparisons

Comparison	OLS, all sectors, mono-establishments				
	A	D	S	R^2	obs.
	(1)	(2)	(3)	(4)	(5)
Cities with pop. > 200,000 vs. pop.< 200,000	0.09 (0.00) *	1.22 (0.01) *	0.00 (0.00)	1.00	139,143
Employment areas above vs. below median density	0.09 (0.00) *	1.21 (0.01) *	0.00 (0.00)	1.00	139,143
Paris vs. cities with pop. 1–2 million	0.12 (0.00) *	1.13 (0.02) *	0.00 (0.00)	0.98	47,480
Cities with pop. 1–2 million vs. pop. 200,000–1 million	0.04 (0.00) *	1.05 (0.02) *	0.00 (0.00)	0.95	37,443
Cities with pop. 200,000–1 million vs. pop. < 200,000	0.01 (0.00) *	1.09 (0.01) *	0.00 (0.00)	0.97	91,666
Paris vs. Lyon (pop. 10,381,376 vs. 1,529,824)	0.09 (0.01) *	1.13 (0.03) *	0.00 (0.01)	0.92	41,738
Lyon vs. Grenoble (pop. 1,529,824 vs. 486,022)	0.03 (0.01) *	1.05 (0.05)	0.00 (0.01)	0.62	6,925
Grenoble vs. Troyes (pop. 486,022 vs. 168,605)	0.05 (0.02) *	1.11 (0.09)	0.01 (0.02)	0.86	1,943

detailed, sector by sector, results for French employment areas to compare with those of Table 2. The results are very close. We conclude that using two groups of cities according to population size or two groups of employment areas according to employment density yields very similar results.

The next three lines of Table 4 return to the urban zoning but divide French cities and towns into four groups instead of just two. They are Paris, other cities with population above one million (Lyon, Marseille, and Lille), cities with population between 200,000 and one million, and the rest of the country. Starting with the estimate of A, average TFP for establishments located in Paris is about 12 percent higher than for establishments in other cities with a population above one million (the population of Paris is between six and ten times larger). There is also a sizeable gap between these three other large cities with population above one million and cities with population between 200,000 and one million (the estimate for A is about one third as large but then so is roughly the gap in terms of average population). Finally the gap between the last two groups is small but nonetheless statistically significant. Turning to the dilation parameter D, this remains significantly above unity when considering this more detailed grouping of cities. The distribution of firms' log productivities appears more dilated in larger cities than in smaller cities when considering any two consecutive groups of cities, capturing once again the larger boost for the most productive firms. Finally, regarding the truncation parameter S, our previous results are confirmed with no difference in market selection between any two consecutive groups of cities.

Grouping cities, as we have done so far, is useful because it ensures that we have enough observations to estimate parameters accurately and reduces the impact of idiosyncrasies associated with any particular city. Nevertheless, we now report the in last three lines of Table 4 results for

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), Grenoble (about half a million), Troyes (a smaller city, with a population below 200,000). Although the number of observations becomes small for the comparison between Grenobles and Troyes, the estimate of A remains significant. A trebling of population between Troyes and Grenoble or between Grenoble and Lyon is associated with a 3 to 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 would be expected in light of previous results, there are no differences in the strength of selection. Note also that the fit deteriorates in this last part of Table 4, since we must work with a much smaller number of observations for the last two rows comparing individual cities.

8. Discussion

The consequences of unobserved prices

As is often the case in the estimation of production functions, we do not observe prices and estimate TFP by studying how much value (instead of physical output) an establishment can produce with given inputs. Even if prices were observed, it would be unclear whether higher prices reflect higher price markups or higher quality products, that is the ability of firms to produce more value out of the same inputs.[21]

While we cannot solve this missing price problem (and the related quality issue), we now show that, to the extent it may affect our estimates, it cannot be driving our key results. In fact, if anything, it would work against us by leading us to underestimate the effects of agglomeration (greater shift and dilation in larger cities). The inability to observe prices implies that our estimated log productivities do not capture the physical quantity of output relative to inputs, but instead the value of output relative to inputs. Expressed in terms of our model, we do not measure ϕ, as given by equation (50), but instead

$$\psi = \ln\left(\frac{pQ}{l}\right) = \ln(p) + A_i - D_i \ln(h) = \ln(p) + \phi \,. \tag{63}$$

Thus, by not taking prices out, we are shifting log productivities rightwards by the value of log prices, $\ln(p)$. The problem is that log prices are systematically related both to city size (through $\bar{h}$) and to individual productivity (through h). Recall that, by equation (6), prices are given by $p = \frac{1}{2}(h + \bar{h})$.

---

[21]To solve these two problems, a first possibility is to focus on homogeneous goods for which quantities are directly observed (Syverson, 2004, Collard-Wexler, 2007, Foster, Haltiwanger, and Syverson, 2008). An alternative is to focus on specific industries with localised markets for which direct measures of quality are available, like newspapers and restaurants (Berry and Waldfogel, 2006). These two solutions can only be applied to a small number of industries. The last alternative in the literature is to consider detailed product-level information, including prices, to recover the price markup of firms and back up their 'true' productivity (De Loecker, 2007). With this approach, 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 (e.g., firms selling to consumers with a constant elasticity of substitution across products).

In terms of the relationship with city size,

$$\frac{\partial \ln(p)}{\partial \bar{h}} = \frac{1}{h + \bar{h}} > 0 .$$ (64)

If $\bar{h}$ differs across cities, then by looking at ψ instead of ϕ we are overestimating log productivities for every for h, but we are doing so by more in smaller cities (where $\bar{h}$ is then larger). Hence one consequence of not observing prices is that we may 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,

$$\frac{\partial^2 \ln(p)}{\partial \bar{h} \, \partial h} = -\frac{1}{(h + \bar{h})^2} < 0 .$$ (65)

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

To sum up, if $\bar{h}$ differs across cities (which necessarily implies $S \neq 0$), then by not observing prices we will underestimate the consequences of agglomeration (the shift captured by A and the dilation captured by D). However, recall that our results show that $S = 0$ for all sectors combined and for nearly all individual sectors. Note also that not observing prices does not affect the estimation of S (the value of log TFP at which each distribution might be left-truncated changes, but the share S of establishments in the small city distribution that are truncated out of the large city distribution does not). Thus, our finding that $S = 0$ also implies that estimating TFP through value added does not bias our estimates of A and D.

A comparison with approaches based on summary statistics

A key innovation of our approach is that we simultaneously consider agglomeration economies and market selection as possible causes of the productivity advantages of large cities using the entire distribution of firm log productivities. While this makes a comparison of our findings with those in the literature difficult, we now briefly relate our results to previous contributions studying either agglomeration economies or market selection alone 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 total population or employment density. More recent studies have paid particular attention to addressing two potential problems for identifying agglomeration economies with that approach. First, it is possible that the higher productivity of individual establishments is a cause rather than a consequence of larger or denser cities. This could happen because more productive cities tend to attract more workers and firms and, as a result, become larger and denser. It is also possible that size and productivity are simultaneously determined. This could happen if there is some local characteristic that is correlated with both size and productivity. 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. A second identification issue is that more productive workers may sort into denser areas. This could happen if skilled workers have a stronger preference for the amenities typically found in larger cities or if larger cities provide greater productive benefits for more skilled workers. 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 size or density (Glaeser and Maré, 2001, Combes, Duranton, and Gobillon, 2008). This issue of endogenous labour quality turns out to be substantially more important in practice than endogenous labour quantity. 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.

We can compare the magnitude of agglomeration economies that we find with that of earlier studies looking at agglomeration alone. We do this by turning our estimate of A (the common shift in productivity of large city establishments relative to their small city counterparts) into an elasticity of average TFP with respect to density. An average employee in a French city with population above 200,000 works in a municipality 10.5 times as dense as an average person in the rest of the country. This ratio of 10.5 implies that our estimate of $A = 0.09$ for all sectors combined in Table 2 is equivalent to an elasticity of TFP with respect to employment density of 0.038. Using the same French data as we do but quantifying agglomeration economies by the usual approach of regressing mean TFP for French employment areas on employment density in those areas, Combes *et al.* (2009) find an elasticity of 0.035. More broadly this estimate of 0.038 is within the usual range of between 0.02 and 0.10. The main conceptual difference of our approach with respect to this literature on agglomeration is that we take into account that the higher average productivity in large cities could be caused by stronger selection as well as by stronger agglomeration economies.

Turning to market selection, existing approaches are harder to compare to ours. Our model shares with other models relating selection to market size, such as Syverson (2004) and Melitz and Ottaviano (2008), the prediction that tougher competition leads to a left truncation of the distribution of productivity in large cities relative to small cities. Unfortunately, detecting left truncation on the basis of summary statistics such as the mean or variance of firms' productivity is not straightforward. Greater left truncation increases average productivity, but so does any model of 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 productivities. We note that this result depends crucially on distributional assumptions.[22] Furthermore, it is possible that the strength of both selection and agglomeration increase with city size 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

---

[22]The result that the variance of productivity increases 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).

based on dispersion measures alone very difficult. A key difference of our approach is that we consider simultaneously selection and agglomeration and look at all the moments 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 very particular characteristics. We look instead at a broad cross-section of sectors.[23]

Given these differences with existing approaches, a detailed comparison of results would not be informative. Instead we now 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 so as to equalise mean productivity in small and large cities with $A = 0$ and $D = 1$. We find that to explain a difference in mean log TFP of 0.09 between cities with population above and below 200,000 (as we have here for all sectors), S should be equal to 0.14. 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 a full order of magnitude larger than our current estimates.

9. Concluding comments

To assess the relative importance of local agglomeration effects and market selection to explain the higher productivity of firms in 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 firms' productivity distribution while stronger agglomeration right-shifts and dilates it. A similar prediction would emerge from a much broader class of models nesting agglomeration and selection provided the underlying distribution of the firms' productivities 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 agglomeration effects relative to selection.

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. When applied to distributions of firms' log productivities, this quantile approach is fully consistent with our theoretical framework.

Our main finding is that productivity differences across urban areas in France are mostly explained by agglomeration. The distribution of firms' log productivities in large French cities is remarkably well described by taking the distribution of firms' productivities in small French cities, dilating it, and shifting it to the right. This holds for the productivity distributions of firms across all sectors as well as most two-digit sectors when considered individually. It also holds when comparing cities with population above and below 200,000, more finely subgroups of cities,

---

[23]The approach developed in Del Gatto *et al.* (2006) and 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.

and even individual cities. This finding is also robust to the choice of zoning since taking French employment areas instead of French urban yields similar results. Our bottom line is that, relative to the rest of the country, the distribution of firms log productivities in cities with population above 200,000 is shifted to the right by of 0.09 and dilated by a factor of 1.22. Firms in large cities are thus on average about 9 percent more productive than in small cities. Because of dilation, this productivity advantage is only of 5 percent for firms at the bottom quartile and 14 percent for firms at the top quartile. On the other hand we find no difference between small and large cities 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 (Duranton 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 firms' log productivity distributions 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. 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 very different spatial scale, we also suspect that for many consumer services selection could be very potent at a very 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 more than a new machine-tool producer will affect other machine-tool producers in the same city.

Appendix A. Proof of Lemma 1 and Lemma 1′

We now prove Lemma 1′. The proof of Lemma 1 is a particular case of this for $D_i = D_j = 1$ and hence $D = 1$. Consider first the case $S_i > S_j$. We apply the change of variables $\phi \rightarrow \frac{\phi - A}{D}$, which

turns equation (53) into

$$F_j\left(\frac{\phi - A}{D}\right) = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_j}{1 - S_j}\right\}. \tag{A1}$$

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

$$\frac{F_j\left(\frac{\phi - A}{D}\right) - S}{1 - S} = \max\left\{\frac{-S}{1 - S}, \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_i}{1 - S_i}\right\}. \tag{A2}$$

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

$$\max\left\{0, \frac{F_j\left(\frac{\phi - A}{D}\right) - S}{1 - S}\right\} = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_i}{1 - S_i}\right\} = F_i(\phi). \tag{A3}$$

Consider now the case $S_i < S_j$. We apply the change of variables $\phi \to D\phi + A$, which turns equation (52) into

$$F_i(D\phi + A) = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_j}{D_j}\right) - S_i}{1 - S_i}\right\}. \tag{A4}$$

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

$$\frac{F_i(D\phi + A) - \frac{-S}{1-S}}{1 - \frac{-S}{1-S}} = \max\left\{S, \frac{\tilde{F}\left(\frac{\phi - A_j}{D_j}\right) - S_j}{1 - S_j}\right\}. \tag{A5}$$

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

$$\max\left\{0, \frac{F_i(D\phi + A) - \frac{-S}{1-S}}{1 - \frac{-S}{1-S}}\right\} = \max\left\{0, \frac{\tilde{F}\left(\frac{\phi - A_j}{D_j}\right) - S_j}{1 - S_j}\right\} = F_j(\phi). \tag{A6}$$

$\square$

Appendix B. Implementation

In this appendix, we explain how we compute the minimisation criterium of equation (42), used to estimate the values of the parameters. We do so for our main results, including the dilation parameter D introduced in Section 6. For our baseline results, the implementation is the same with $D = 1$.

First note that the data consist of a set of log productivities in large cities (indexed by i) and in small cities (indexed by j), ranked in ascending order and denoted Φ_i and Φ_j respectively. From these data, for any θ, we need to be able to evaluate $\hat{m}_\theta(u)$ and $\hat{\tilde{m}}_\theta(u)$ at any ranks $u \in [0,1]$ to compute $M(\theta) = \int_0^1 [\hat{m}_\theta(u)]^2 du + \int_0^1 [\hat{\tilde{m}}_\theta(u)]^2 du$. For that purpose, we construct some estimators

37

$\hat{\lambda}_i$ and $\hat{\lambda}_j$ of the quantiles $\lambda_i(u)$ and $\lambda_i(u)$. Focusing on large cities (replace i with j for small cities), we start from the set of log productivities

$$\Phi_i = [\phi_i(0), \ldots, \phi_i(E_i - 1)]' , \tag{A7}$$

where E_i is the number of establishments in i and $\phi_i(0) < \ldots < \phi_i(E_i - 1)$. We can construct the sample quantiles at the observed ranks as $\hat{\lambda}_i\left(\frac{k}{E_i}\right) = \phi_i(k)$ for $k \in \{0, \ldots, E_i - 1\}$. For any other rank $u \in]0,1[$, the estimators of the quantiles are recovered by linear interpolation:

$$\hat{\lambda}_i(u) = (k_i^* + 1 - uE_i)\,\hat{\lambda}_i\left(\frac{k_i^*}{E_i}\right) + (uE_i - k_i^*)\,\hat{\lambda}_i\left(\frac{k_i^* + 1}{E_i}\right) , \tag{A8}$$

where $k_i^* = \lfloor uE_i \rfloor$ and $\lfloor . \rfloor$ denotes the integer part. From equation (A8) and the corresponding expression for j, we can use the empirical counterparts of equations (61) and (62),

$$\hat{m}_\theta(u) = \hat{\lambda}_i\left(r_S(u)\right) - D\hat{\lambda}_j\left(S + (1 - S)r_S(u)\right) - A , \tag{A9}$$

$$\hat{\tilde{m}}_\theta(u) = \hat{\lambda}_j\left(\tilde{r}_S(u)\right) - \frac{1}{D}\hat{\lambda}_i\left(\frac{\tilde{r}_S(u) - S}{1 - S}\right) + \frac{A}{D} , \tag{A10}$$

to compute $\hat{m}_\theta(u)$ and $\hat{\tilde{m}}_\theta(u)$ at any rank u and for any θ. We then consider $K = 1001$ ranks evenly distributed over the interval $[0,1]$. These ranks are denoted u_k, $k \in \{0, \ldots, K\}$, with $u_0 = 0$ and $u_K = 1$. We approximate the two subcriteria using the formulas:

$$\int_0^1 [\hat{m}_\theta(u)]^2 \mathrm{d}u \approx \frac{1}{2} \sum_{k=1}^K \left\{ [\hat{m}_\theta(u_k)]^2 + [\hat{m}_\theta(u_{k-1})]^2 \right\} (u_k - u_{k-1}) , \tag{A11}$$

$$\int_0^1 [\hat{\tilde{m}}_\theta(u)]^2 \mathrm{d}u \approx \frac{1}{2} \sum_{k=1}^K \left\{ [\hat{\tilde{m}}_\theta(u_k)]^2 + [\hat{\tilde{m}}_\theta(u_{k-1})]^2 \right\} (u_k - u_{k-1}) . \tag{A12}$$

The estimated parameters $\hat{\theta}$ are those which minimise the sum of these two quantities.

Appendix C. Implementation of alternative approaches to productivity

Olley-Pakes

In this appendix, we present three alternative approaches to TFP estimation. The first is the methodology proposed by Olley and Pakes (1996) to account for the endogeneity of production factors when estimating the parameters of equation (46). These authors consider that the residual ϕ_t can be decomposed into an unobserved factor φ_t which is potentially correlated with labour and capital, and an uncorrelated error term η_t such that: $\phi_t = \varphi_t + \eta_t$. They suppose that the unobserved factor φ_t can be rewritten as its projection on its lag and an innovation: $\varphi_t = \kappa(\varphi_{t-1}) + \xi_t$. They also make the crucial assumption that capital investment at time t depends on the capital stock and the unobserved factor φ_t: $I_t = i_t(k_t, \varphi_t)$. The function i_t is supposed to be strictly increasing in the unobserved factor. It can be inverted such that: $\varphi_t = f_t(k_t, I_t)$. Equation (46) can then be rewritten as:

$$\ln(V_t) = \beta_2 \ln(l_t) + \sum_{s=1}^3 \sigma_s l_{s,t} + \Psi_t(k_t, I_t) + \eta_t , \tag{A13}$$

where the auxiliary function Ψ_t is defined as

$$\Psi_t(k_t, I_t) = \beta_{0,t} + \beta_1 \ln(k_t) + f_t(k_t, I_t) \ . \tag{A14}$$

Equation (A13) can be estimated with OLS after $\Psi_t(k_t, I_t)$ has been replaced with a third-order polynomial crossing k_t, I_t and year dummies. This allows to recover some estimators of the labour and skill share coefficients ($\hat{\beta}_2$ and $\hat{\sigma}_s$), as well as the auxiliary function ($\hat{\Psi}_t$). It is then possible to construct the variable

$$v_t = \ln(V_t) - \hat{\beta}_2 \ln(l_t) - \sum_{s=1}^{3} \hat{\sigma}_s l_{s,t} \ . \tag{A15}$$

From equation (A14), the lagged value of the unobserved factor φ_{t-1} can be approximated by $\hat{\Psi}_{t-1}(k_{t-1}, I_{t-1}) - \beta_{0,t-1} - \beta_1 \ln(k_{t-1})$. Using equations (A13), (A14), (A15), and the projection of the unobserved factor on its lag, the value-added equation then becomes:

$$v_t = \beta_{0,t} + \beta_1 \ln(k_t) + \kappa \left(\hat{\Psi}_{t-1}(k_{t-1}, I_{t-1}) - \beta_{0,t-1} - \beta_1 \ln(k_{t-1}) \right) + \vartheta_t \ , \tag{A16}$$

where ϑ_t is a random error. The function $\kappa(.)$ is approximated by a third-order polynomial and equation (A16) is estimated with non-linear least squares. We thus recover some estimators of the year dummies ($\hat{\beta}_{0,t}$) and the capital coefficients ($\hat{\beta}_1$). An estimator of ϕ_t is then given by $\hat{\phi}_t = v_t - \hat{\beta}_{0,t} - \hat{\beta}_1 \ln(k_t)$.

   Although the Olley-Pakes method allows us to control for simultaneity, it has some drawbacks. In particular, we need to construct investment from the data: $I_t = k_t - k_{t-1}$. Since investment enters lagged into equation (A16), we must observe firms for at least three consecutive years to compute their TFP with this method. Other observations must be dropped. Furthermore, the investment equation $I_t = i_t(k_t, \varphi_t)$ can be inverted only if $I_t > 0$. Hence, we can keep only observations for which $I_t > 0$. This double selection may introduce a bias, for instance, if (i) there is greater 'churning' (i.e. entry and exits) in denser areas, and (ii) age and investment affect productivity positively. Then, more establishments with a low productivity may be dropped in high density areas. In turn, this may increase the measured difference in local productivity between areas of low and high density. Re-estimating OLS TFP on the same sample of firms used for Olley-Pakes shows that this is, fortunately, not the case on French data.

Levinsohn-Petrin

We also implement the approach proposed by Levinsohn and Petrin (2003). Its main difference with Olley and Pakes (1996) is that the quantity of inputs is used to account for the unobservables instead of investment. The unobserved factor is then rewritten as $\varphi_t = f_t(k_t, Ic_t)$ where Ic_t is the consumption of inputs. Otherwise, the estimation procedure remains the same. However, we lose fewer observations since the use of materials instead of investment means we need to observe firms for two consecutive years instead of three.

Cost shares

Alternatively, a TFP measure can be constructed using cost shares as estimates of the labour and capital coefficients in equation (46). The costs of labour and capital were evaluated by Boutin

and Quantin (2006) for each cell defined by the 3-digit industry, the year, and the number of firm employees (less than 5, 5–20, 20–50, 100, more than 100). The share of capital (resp. labour) in these costs is denoted $\hat{\beta}_{1,t}^s$ (resp. $\hat{\beta}_{2,t}^s$). Implicitly, we assume that returns to scale equal one as we have: $\hat{\beta}_{1,t}^s + \hat{\beta}_{2,t}^s = 1$. The predicted value-added based on capital and labour is $\ln V_t^p = \hat{\beta}_{1,t}^s \ln(k_t) + \hat{\beta}_{2,t}^s \ln(l_t)$. The following specification can then be estimated with OLS:

$$\ln(V_t) - ln V_t^p = \beta_{0,t} + \sum_{s=1}^3 \sigma_s l_{s,t} + \widetilde{\phi}_t \tag{A17}$$

Denoting $\hat{\beta}_{0,t}^s$ and $\hat{\sigma}_s^s$ the estimated coefficients, the TFP measure is given by:

$$\hat{\widetilde{\phi}}_t = \ln(V_t) - ln V_t^p - \hat{\beta}_{0,t}^s - \sum_{s=1}^3 \hat{\sigma}_s^s l_{s,t} \tag{A18}$$

For all methods, the TFP of a firm is the firm-level average of yearly TFP over the period 1994–2002. The TFP estimates we recover with these four approaches are highly correlated. The correlation between OLS TFP and Olley-Pakes TFP is 0.73. The correlation between OLS TFP and Levinsohn-Petrin TFP is 0.85. The correlation OLS TFP and cost-shares TFP is 0.93. Unsurprisingly, these alternative methods to estimate TFP give results which are qualitatively similar for A, D, and S at the sector level. These results are available upon request.

Appendix D. Estimations for employment areas

Table 5: Employment areas above vs. below median density

| Sector | OLS, mono-establishments | | | | obs. |
| | A | D | S | R^2 | |
	(1)	(2)	(3)	(4)	(5)
Food, beverages, tobacco	0.07 (0.00) *	0.94 (0.02) *	0.01 (0.00) *	0.97	22,049
Apparel, leather	0.04 (0.01) *	1.35 (0.05) *	0.01 (0.01)	0.99	5,804
Publishing, printing, recorded media	0.17 (0.01) *	1.23 (0.05) *	0.00 (0.00)	0.98	9,236
Pharmaceuticals, perfumes, soap	0.06 (0.06)	1.15 (0.13)	-0.01 (0.06)	0.86	1,069
Domestic appliances, furniture	0.13 (0.01) *	1.20 (0.04) *	0.01 (0.01) *	0.98	6,362
Motor vehicles	0.09 (0.02) *	1.19 (0.13) *	0.01 (0.02)	0.85	1,442
Ships, aircraft, railroad equipment	0.09 (0.03) *	1.10 (0.17)	0.00 (0.02)	0.85	1,016
Machinery	0.09 (0.01) *	1.06 (0.02) *	0.00 (0.00)	0.98	14,738
Electric and electronic equipment	0.08 (0.01) *	0.98 (0.04)	0.00 (0.00)	0.96	5,749
Building materials, glass products	0.07 (0.01) *	1.10 (0.05)	0.00 (0.01)	0.94	3,196
Textiles	0.06 (0.01) *	1.11 (0.07) *	0.01 (0.01)	0.94	3,363
Wood, paper	0.09 (0.01) *	1.13 (0.05) *	0.00 (0.01)	0.98	5,872
Chemicals, rubber, plastics	0.08 (0.01) *	1.11 (0.04) *	0.01 (0.01)	0.96	5,335
Basic metals, metal products	0.07 (0.01) *	1.06 (0.02) *	0.00 (0.00)	1.00	14,305
Electric and electronic components	0.08 (0.01) *	0.99 (0.06)	-0.01 (0.01)	0.94	2,579
Consultancy, advertising, business services	0.19 (0.02) *	1.06 (0.03)	-0.01 (0.03) *	0.96	37,041
All sectors	0.09 (0.00) *	1.21 (0.01) *	0.00 (0.00)	1.00	139,143

References

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.

Berry, Steven and Joel Waldfogel. 2006. Product quality and market size. Processed, Yale University.

Boutin, Xavier and Simon Quantin. 2006. Une méthodologie d'évaluation comptable du coût du capital des entreprises françaises: 1984–2002. Working Paper G2006–09, INSEE-DESE.

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.

Ciccone, Antonio and Robert E. Hall. 1996. Productivity and the density of economic activity. *American Economic Review* 86(1):54–70.

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.

Collard-Wexler, Allan. 2007. Productivity dispersion and plant selection in the ready-mix concrete industry. Processed, New York University.

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, Laurent Gobillon, and Sébastien Roux. 2009. Estimating agglomeration effects with history, geology, and worker fixed-effects. In Edward L. Glaeser (ed.) *The Economics of Agglomeration*. Cambridge, MA: National Bureau of Economic Research, forthcoming.

De Loecker, Jan. 2007. Product differentiation, multi-product firms and estimating the impact of trade liberalization on productivity. Processed, Princeton University.

Del Gatto, Massimo, Giordano Mion, and Gianmarco I.P. Ottaviano. 2006. Trade integration, firm selection and thecosts of non-Europe. Discussion Paper 5730, Centre for Economic Policy Research.

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.

Dixit, Avinash K. 1979. A model of duopoly suggesting a theory of entry barriers. *Bell Journal of Economics* 10(1):20–32.

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.

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. 2008. Quantile-based inference of parametric transformations between two distributions. Processed, CREST-INSEE.

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.

Hopenhayn, Hugo. 1992. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica* 60(5):1127–1150.

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. 2007. Improved market access to foreign markets raises plant-level productivity... for some plants. Working Paper 13297, National Bureau of Economic Research.

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.

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:forthcoming.

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.

Ottaviano, Gianmarco I. P., Takatoshi Tabuchi, and Jacques-Franccois Thisse. 2002. Agglomeration and trade revisited. *International Economic Review* 43(2):409–436.

Pavcnik, Nina. 2002. Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants. *Review of Economic Studies* 69(1):245–276.

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.

Syverson, Chad. 2004. Market structure and productivity: A concrete example. *Journal of Political Economy* 112(6):1181–1222.

Vives, Xavier. 1990. Trade association disclosure rules, incentives to share information, and welfare. *Rand Journal of Economics* 21(3):409–430.

Wheaton, William C. and Mark J. Lewis. 2002. Urban wages and labor market agglomeration. *Journal of Urban Economics* 51(3):542–562.