INSTITUT NATIONAL DE LA STATISTIQUE ET DES ETUDES ECONOMIQUES
Série des Documents de Travail du CREST
(Centre de Recherche en Economie et Statistique)

n° 2010-52

The Identification of Agglomeration Economies

P.-Ph. COMBES[1]
G. DURANTON[2]
L. GOBILLON[3]

Les documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.

Working papers do not reflect the position of INSEE but only the views of the authors.

[1] GREQAM, Université d'Aix-Marseille.
[2] Université de Toronto, Canada.
[3] Institut National d'Etudes Démographiques, PSE et CREST.

The Identification of Agglomeration Economies

Pierre-Philippe Combes[*,†]
Greqam, *Aix-Marseille University*

Gilles Duranton[*,‡]
University of Toronto

Laurent Gobillon[*,§]
Institut National d'Etudes Démographiques, PSE and CREST

April 2010

Abstract: Measures of urban productivity are typically positively associated with city population. But is this relationship causal? We discuss the main sources of bias in the proper identification of agglomeration effects. We also assess a variety of solutions that have been proposed in the literature to deal with them.

Key words: agglomeration economies, instrumental variables, structural estimation, natural experiments

jel classification: r12, r23

[*] Many thanks to Thierry Magnac and Diego Puga for a number of useful conversations. Financial support from the Centre National de la Recherche Scientifique (Combes) and from the Canadian Social Science and Humanities Research Council (Duranton) is gratefully acknowledged.

[†] 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/). Also affiliated with the Centre for Economic Policy Research, International Trade and Regional Economics programme.

[‡] 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). Also affiliated with the Centre for Economic Policy Research, International Trade and Regional Economics programme.

[§] 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/). Also affiliated with the Centre for Economic Policy Research, International Trade and Regional Economics programme.

1. Introduction

The economic approach to cities argues that cities result from a tradeoff between urban costs and urban benefits, both of which increase with urban scale. If there were only costs to cities, we would all disperse. If there were only benefits, we would all agglomerate in one city. To put this cornerstone idea about cities on a safe empirical footing, research has gone beyond this 'out-of-necessity' argument and has attempted to measure the benefits from cities. It has done so mostly by focusing on how urban scale affects various measures of urban productivity. This paper discusses the econometric challenges associated with this endeavour and how successful it has been.

More precisely, the agglomeration economies vs. urban costs approach to cities rests on four claims: (i) there are urban costs; (ii) there are agglomeration economies; (iii) agglomeration economies dominate for small population sizes whereas urban costs dominate for large sizes; (iv) there is some mobility across cities (or between cities and their rural hinterland) so that the size of a city depends on what existing residents might get elsewhere.[1] Empirical research has taken (i) to be mostly obvious. Land is in finite quantity. So is the roadway. Bringing in more people should raise urban costs. The focus of extant research has instead been on (ii), the measurement of the benefits from agglomeration. This is an obvious prerequisite to dealing with (iii) and (iv).

The proper identification of agglomeration benefits is also important for a range of other reasons. Many local public policy initiatives attempt to foster agglomeration economies by building clusters, attracting 'talent', or luring large industrial facilities. The benefits to be expected from such policies are in direct relation with the intensity of agglomeration economies. It is also the case that a full cost-benefit analysis of most urban infrastructure projects requires knowing about agglomeration effects. For instance, a new urban highway or a new transit line might affect agglomeration both directly by easing interactions within the city and indirectly through population and employment growth.

The scope of this paper is purposefully narrow and our review intends to be selective. We are interested in measuring the elasticity of various measures of urban productivity, mainly wages, with respect to urban scale and we examine the econometric pitfalls associated with this enterprise. We do not try to distinguish between possible sources of agglomeration economies. We do not broaden the discussion to related topics such as urban growth either. Melo, Graham, and Noland (2009) propose a meta-analysis of all published results on the topic. The sources of agglomeration economies are discussed by Rosenthal and Strange (2004) and Puga (2010). Henderson (2007) proposes a broader discussion of agglomeration effects.

In the following, section 2 derives a simple specification that has been often used in the literature. The sources of bias are discussed in section 3. Our assessment of the various solutions proposed in the literature to overcome these biases is in section 4 before our conclusion.

[1] Urban costs are at the heart of the monocentric model of cities developed by Alonso (1964), Mills (1967), and Muth (1969), the founding model of urban economics. The integration of costs and benefits of cities in a model of an entire urban system is by Henderson (1974).

2. The model

We start with a simple model. It allows us to describe key identification issues in the estimation of agglomeration economies. Firm j located in city c operates under constant returns. Its output y_j is

$$y_j = A_j k_j^{\theta} l_j^{1-\theta} ,$$

where A_j is technology, k_j capital, and l_j labour measured in efficiency units. In turn, labour is given by

$$l_j = \sum_i e_{ij} \ell_i,$$

where ℓ_i is the number of hours worked by worker i and e_{ij} measures the efficiency of this worker. With competitive markets for final goods and inputs, the first-order conditions for profit maximisation reduce to

$$w_{ij} = \Phi_j \, e_{ij} \, , \tag{1}$$

where w_{ij} is the hourly wage of worker i and Φ_j is a wage shifter for firm j defined as

$$\Phi_j \equiv (1-\theta)\theta^{\theta/(1-\theta)} \left(\frac{p_j A_j}{r_j^{\theta}} \right)^{1/(1-\theta)} . \tag{2}$$

with p_j the revenue per unit sold (net of intermediate consumptions and trade costs, if any, born on exported units) and r_j is the cost of capital, k_j.

Equations (1) and (2) summarise the key effects. Wages are high in cities that benefit from pure externalities which increase either technology, A_j, or labour efficiency, e_{ij}. Wages are also high when market access is good as it implies high prices net of trade costs, p_j. Finally, wages are also higher when the cost of capital, r_j, is low. More generally, if we think more broadly about k_j and have it represent non-labour inputs, r_j will also depend on the price of land and the cost of intermediates used in production. All this implies that urban scale may raise wages through a variety of channels: better technology (A_j), higher labour efficiency (e_{ij}), higher prices (p_j), and lower costs of other factors (r_j).

While we return to most of them below, a number of issues must be kept in mind before we proceed. First, because we cannot track technology independently from input prices or the costs of other factors, we can only estimate a 'net' effect of agglomeration That is, we assess the effect of urban scale on the wage shifter Φ_j in equation (1). Second, the assumption of competitive markets for inputs and final goods may raise some concerns. Given our objective, these concerns are not major. To understand why, note first that if prices were a fixed markup on marginal cost (or if wages were a fixed fraction of the marginal product of labour), equation (1) would be barely affected – only the constants would differ. However, price markups may change with urban scale. In this case, any pro-competitive effects associated with lower price markups in larger cities will be estimated as part of the 'agglomeration' effects. Third, we focus on wages. A legitimate alternative is to estimate total factor productivity (TFP) and assess how it is affected by urban scale. We believe that, in practice, this alternative approach has only one minor advantage and a significant drawback. With a reasonable measure of capital, one can condition out differences in capital intensity. This may remove a confounding factor.[2] On the other hand, current approaches to TFP are weak at dealing with differences in labour quality, a fundamental issue in the estimation of agglomeration effects. Fourth, the model accounts for differences in workers' individual efficiency, e, in a limited way. Workers are assumed to be perfect substitutes after their labour is expressed in efficiency units. In practice, we expect some substitution across skill groups. This specification issue remains unresolved.[3] Finally, our model is purposefully rudimentary but we note that models of agglomeration often lead to a similar reduced form where wages increase with urban scale (Duranton and Puga, 2004).

Equation (1) can be rewritten as $\log w_{ij} = \log \Phi_j + \log e_{ij}$. For now, we assume that all firms in city c have the same technology and face the same input prices and trade costs for their output so that $\log w_i = \log \Phi_c + \log e_{ic}$ where the wage shifter Φ_c is a function of urban scale and worker i's efficiency e depends on some permanent characteristics of that worker and some city specific shock for that worker. To measure urban scale, we use density (population or employment per unit of land-area) and assume a log-linear effect. For worker i, when located in city c, we end up with:

$$\log w_{ic} = \alpha \log den_c + \eta_c + u_i + \varepsilon_{ic} \,, \tag{3}$$

where den_c is density in city c, η_c is a city-effect aimed at capturing characteristics of city c including any natural advantage, u_i is a worker effect, and ε_{ic} is a shock specific to worker i in city c.

In the data, we do not observe the worker and city effects u_i and η_c, which typically enter the error term. We do not observe the whole set of possible wages that worker i would get in all cities either. We only observe the wage in the city $c(i)$ in which she is located. Therefore the specification that is often estimated is not (3) but instead:

$$\log w_{ic(i)} = \alpha \log den_{c(i)} + \eta_{c(i)} + u_i + \varepsilon_{ic(i)} \,, \tag{4}$$

or something more aggregated when only city averages are observed. Usually, the wage is denoted w_i instead of $w_{ic(i)}$, the shock ε_i instead of $\varepsilon_{ic(i)}$, city density den_c instead of $den_{c(i)}$ and so on. This somewhat hides that location choices can be endogenous, the reason why we prefer our notation. This said, in regression (4), the main parameter of interest is α, the elasticity of wages with respect to urban density.

3. The sources of estimation bias

An OLS estimate for α in regression (4) is unbiased under the assumption that the covariance:

$$\begin{aligned}\mathbb{C}\text{ov}\left(\log den_{c(i)}, \eta_{c(i)} + u_i + \varepsilon_{ic(i)}\right) &= \mathbb{C}\text{ov}\left(\log den_{c(i)}, \eta_{c(i)}\right) \\ &+ \mathbb{C}\text{ov}\left(\log den_{c(i)}, u_i\right) + \mathbb{C}\text{ov}\left(\log den_{c(i)}, \varepsilon_{ic(i)}\right)\end{aligned} \tag{5}$$

[2] However, other factors of production such as land remain typically unobserved. Prices are also unobserved (or if they are it is unclear what they mean given quality differences). Hence, we do not expect standard ways to estimate TFP to yield credible estimates for A_j. They are more likely to provide an estimate of Φ_j.

[3] Ciccone and Peri (2006) consider the effects of substitution across skill groups in a related problem, the estimation of human capital externalities.

is zero. When it differs from zero, any of the three covariances on the right-hand side of (5) is a source of bias of the OLS estimator. We discuss these three covariance terms in turn.

31 *Endogenous quantity of labour*

Ignoring the endogeneity of location for now, the first term on the right-hand side of equation (5), which reduces to $\mathbb{C}\text{ov}\left(\log den_c, \eta_c\right)$, is different from zero when density, den_c, is correlated with city fixed effects, η_c. A first possibility would be that, absent any location choice by workers, cities with a natural advantage in production (high η_c) have greater fertility and end up with higher density.

An arguably more serious concern today is that cities with a natural advantage in production attract more workers. Let us ignore individual heterogeneity for now and consider a representative worker whose utility in city c depends on the wage, $\log w_c$, and consumption amenities, μ_c:

$$U_c = U\left(\log\, w_c, \mu_c\right).$$

In this case, workers go to cities that offer high wages and high amenities.[4] That is, density increases with wages and amenities following aggregate location choices. Let us write this relationship log linearly

$$\log den_c = \beta \log w_c + \mu_c. \tag{6}$$

At the same time, simplifying equation (4) to ignore individual heterogeneity implies $\log w_c = \alpha \log den_c + \eta_c$. Then, we can solve for density using this expression and (6) to get

$$\log den_c = \frac{\beta\eta_c + \mu_c}{1 - \alpha\beta}.$$

From this equation, it is easy to obtain:

$$(1 - \alpha\beta)\, \mathbb{C}\text{ov}\left(den_c, \eta_c\right) = \beta \text{var}(\eta_c) + \mathbb{C}\text{ov}\left(\mu_c, \eta_c\right). \tag{7}$$

This expression allows us to discuss in greater detail the circumstances under which the covariance $\mathbb{C}\text{ov}\left(den_c, \eta_c\right)$ is not zero.[5]

The first component of this covariance contains a term in $\text{var}(\eta_c)$. It arises from the dependence of density on the wage, itself a function of natural advantage. Hence any missing local natural advantage that raises wages and makes cities denser will bias the estimation. The upward bias on α depends on the importance of the population feedback, β.

The second component of the covariance in (7) is less obvious. In essence, the variables that enter η_c as natural advantages might be correlated with amenities μ_c and this results in $\mathbb{C}\text{ov}\left(\eta_c, \mu_c\right) \neq 0$. A first reason for this is that some city attributes affect both city productivity (as natural advantage) and utility (as amenities). For instance, research centres co-locate with large

[4] Of course, we also expect urban costs to increase with density. This does not matter for our purpose here except through the channels we mention below.

[5] Relative to a situation where workers are immobile, it must also be kept in mind that the covariance between density and natural advantages is not calculated in the same way since allowing for mobility leads to population changes in all areas.

universities and are found in dense areas. Research centres might increase productivity and thus be part of η_c. At the same time, universities may also be the source of cultural amenities like theatres and restaurants. All this implies $\mathbb{C}\text{ov}\left(\eta_c, \mu_c\right) \neq 0$. Coastal location could be another example. Being on a coast lowers trade costs and is thus a form of natural advantage in our model since it raises the price net of trade cost p_c. At the same time, workers might value coastal locations as amenities.

A second, and potentially even more serious source of correlation between natural advantage and amenities arises through the land market. As made clear by equation (6), amenities attract workers. In turn, the resulting higher density implies higher land prices. At the same time firms also consume land and land prices (through the term r_c) enter the wage shifter Φ_c in equation (2). This linkage implies a lower η_c. Consequently $\mathbb{C}\text{ov}\left(\eta_c, \mu_c\right) \neq 0$. This mechanism is at the core of Roback's (1982) spatial equilibrium framework which is widely used in empirical work on cities.

In sum, we have both a reverse causation and a missing variable problem. The issue of missing variables is particularly severe since it applies to any variable that either affects productivity directly or that affects any other other components of Φ indirectly.

The endogenous quantity of labour bias is made more complicated when we consider location choices at the worker level. Consider worker i with utility U_{ic} when located in city c. Utility for this worker depends on the expected wage $\log \widetilde{w}_{ic} = \mathbb{E}\left(\log w_{ic} | den_c, \eta_c, u_i\right)$ in city c, consumption amenities μ_c, and the characteristics of the workers u_i:

$$U_{ic} = U\left(log\widetilde{w}_{ic}, \mu_c, u_i\right),$$

so that workers will choose their location based on their expected wage and on their unobserved individual characteristics. An additional issue is that the covariance term in (5) are indexed by the city chosen by the worker. This location choice creates a selection effect on the cities on which the covariance, $\mathbb{C}\text{ov}\left(\eta_{c(i)}, \mu_{c(i)}\right)$, is computed. In other words, we have more individual observations for those cities where both η_c and μ_c are high.

32 *Endogenous quality of labour*

Again, let us start the discussion by abstracting from workers' location choice. The second term on the right-hand side of (5), $\mathbb{C}\text{ov}\left(\log den_c, u_i\right)$, is different from zero when the worker effect u_i is a consequence of local density. Better schools and universities in denser cities can make workers born there more productive. In addition, learning at the workplace may be more important in denser cities as suggested by Glaeser's (1999) model. Empirically the importance of local learning is asserted by Glaser and Maré (2001), Baum-Snow and Pavan (2010), or De la Roca and Puga (2010). Faster learning in denser cities also implies $\mathbb{C}\text{ov}\left(\log den_c, u_i\right) \neq 0$.

Endogenous location choices can magnify these effects. Workers with a better learning potential (and this may be correlated with u_i) may choose to go to denser cities where more learning takes place. But even if individual abilities are drawn independently from density, that is if $\mathbb{C}\text{ov}\left(\log den_c, u_i\right) = 0$, we may have $\mathbb{C}\text{ov}\left(\log den_{c(i)}, u_i\right) \neq 0$ because of ability sorting. This can happen because of a complementarity between workers' skills and density (or city size). Such complementarity appears for instance in the model of Behrens, Duranton, and Robert-Nicoud

(2010). Workers with more skills are more likely to become more productive entrepreneurs who stand to gain more from being in denser cities. In turn, this pushes towards the sorting of more skilled individuals in denser cities. Such complementarity between cities and skills is documented empirically by a number of authors including Bacolod, Blum and Strange (2009) and Glaeser and Resseger (2010). Other mechanisms are also possible. For instance, the amenities associated with urban density mentioned above might be more appealing to more skilled workers, etc.

33 *Shocks*

Finally, and ignoring again location choices to begin, the last term on the right-hand side of equation (5), $\mathbb{C}\text{ov}\left(\log den_c, \varepsilon_{ic}\right)$, is different from zero when the realisation of the worker shocks ε_{ic} depends on density. A specific government policy (which would be treated like a shock in a regression) might for instance lead to higher wages for certain groups of workers who are unevenly distributed in space.[6]

Endogenous location choices can also create an extra effect on this covariance. Imagine that workers observe the realisation of their shock in each city and choose their location accordingly. Utility is now given by

$$U_{ic} = U\left(logw_{ic}, \mu_c, u_i\right), \tag{8}$$

that is, the utility of worker i in a given city c depends on the exact wage she gets there, w_{ic}, rather than her expected wage (unlike our previous example). This occurs when the location decision is based on a precise job offer for instance. In that case, we may have $\mathbb{C}\text{ov}\left(\log den_{c(i)}, \varepsilon_{ic(i)}\right) \neq 0$ even though $\mathbb{C}\text{ov}\left(\log den_c, \varepsilon_{ic}\right) = 0$.

34 *An aside on specification*

While not properly speaking a source of bias, it is nonetheless important to clarify at this stage a number of specification issues.

Many models imply that urban scale affects contemporaneous productivity. The level of aggregation for the analysis is thus the city and the regression considers both productivity and urban scale at the same point in time. Beyond the models, the city size elasticity of wages (or productivity) is a natural quantity to look at if one is interested in accounting for spatial wage disparities. Focusing on one elasticity also makes it easier to compare results across studies.

This said, we expect the concentration of economic activity to affect outcomes at various spatial scales with possibly some complex dynamics at play. For instance, the benefits from cities may also include gradual learning by workers. It may then allow them to command a higher wage anywhere. This type of effect needs to be taken into account to assess the overall benefits from cities. However, if the objective here is to estimate the size elasticity of wages, it is only relevant to the extent that it might bias the estimation of the city size elasticity of wage. Furthermore, many interesting interactions take place at a spatial scale finer or broader than cities, neighbourhoods or

[6] One may also imagine that shocks are on average higher in denser cities because, for instance, they generate more wage offers and workers are able to choose the best ones. We tend to think of such effects as being part of what we try to measure as agglomeration effects.

regions. Again, we only worry about these effects to the extent that they are correlated with our key dependent variable, density. Understanding the dynamics of agglomeration effects in cities and agglomeration effects at smaller or larger spatial scales are important endeavours but not part of what we want to estimate here.

We usually prefer to use density as key independent variable to other measures of urban scale such as counts of population or employment. This is because density may be more robust in practice to arbitrary borders for cities.[7] It could also be that the spatial scope of agglomeration effects is geographically limited. Often, density captures 'scale' better than overall population. But between density and population, this need not be one or the other. Arguably, these two variables capture different things and might both be included in the regression. A good case can also be made that aggregate shares of skills also matter (Moretti 2004a) and be included in the regression, etc. Unfortunately, these supplementary independent variables are subject to the same estimation concerns as density. Ideally, we want to be able to exclude them from the analysis while dealing with the estimation problems for density. This may not be always possible. At a practical level, we can only recommend checking what these extra city characteristics do to the estimation of the coefficient of interest, α.

As in much of the literature, we have a log linear specification. Given that both wages and density are right-skewed this is a natural way to estimate the regression. This type of specification also comes straight from theory. But this remains a functional form assumption and nothing guarantees that this is reasonable. Empirically the relationship between urban scale and productivity appears to be log linear in most cases (Au and Henderson, 2006, for China is an exception). Some scrutiny is nonetheless warranted with every new data.

The more important issue is that not all workers and firms may benefit from agglomeration with the same intensity. In addition, not all workers and firms contribute to agglomeration to the same extent. Estimating an 'average' elasticity of wages with respect to size is interesting but knowing about any heterogeneity within this relationship is the logical next step. There are at least three dimensions of interest. The first, and historically the most prominent, is the sectoral dimension. Since at least Henderson (1988), there is a long tradition that looks for agglomeration effects within cities and sectors. A fascinating results is that for mature industries these effects seem to be important whereas it is overall scale that appears to matter more for high-tech industries (e.g., Henderson, Kuncuro, and Turner, 1995). The second dimension is the industrial organisation of firms. Results by Rosenthal and Strange (2003, 2010) suggest that small firms are the main beneficiaries and originators of agglomeration effects. The last one is the skill dimension. Recent evidence by Bacolod, Blum, and Strange (2009) or Glaeser and Resseger (2010) among others suggest that more skilled workers benefit more from being in larger cities.

[7] When using French employment areas, Greater Paris is divided into several units. So is New York when using US counties.

4. Which possible solutions?

We now explore a range of 'solutions' that have been proposed in the literature to deal with the estimations biases described above.

41 *Fixed effects*

Let us start with a 'naive' OLS regression in cross-section:

$$\log w_{ic(i)} = \alpha \log den_{c(i)} + \xi_{ic(i)} \,, \tag{9}$$

where $\xi_{ic(i)} = \eta_{c(i)} + u_i + \varepsilon_{ic(i)}$. This regression is subject to the three main sources of bias examined above. To limit the covariance between $\log den_{c(i)}$ and u_i or $\eta_{c(i)}$, one might consider, as a first step, to introduce worker and city characteristics. They could solve partly the biases that emerge even under exogenous location choices but not those due to the endogeneity of location choices. Some worker characteristics such as age or gender might be added without too much to worry about.[8] Educational achievements are more suspicious since high wage cities might provide more education. Observable city characteristics raise similar concerns since they are likely to be simultaneously determined with wages. Using observables to proxy for the unobservables u_i or $\eta_{c(i)}$ while making sure that the former are exogenous, is a worthwhile element of any reasonable empirical strategy. Nonetheless we cannot expect too much from this first step. Nothing guarantees that 'exogenous' observed city and worker characteristics will capture much of u_i and η_c.

The second step is to introduce fixed effects. This approach was pioneered by Glaeser and Maré (2001). Fixed effects obviously raise the data requirements for the analysis. A single cross-section of average wages by city is no longer enough. We need micro-data and be able to observe workers at least twice. The simplest specification is to impose city and worker fixed effects, η_c and u_i, encompassing any unobserved city or worker characteristic invariant over time, together with some time fixed effects and time-varying worker characteristics added as controls (not shown in the equation):

$$\log w_{ic(it)t} = \alpha \log den_{c(it)t} + \eta_{c(it)} + u_i + \varepsilon_{ic(it)t}, \tag{10}$$

where subscript $c(it)$ refers to the city where worker i is located at date t. This specification raises two issues. City effects are estimated only from the 'movers' (workers who change location between two dates). With only 'stayers' we would not be able to identify city effects separately from worker effects. Estimating from movers means estimating from a possibly highly selected set of individuals. For instance, workers who move to dense cities are arguably those that stand to gain the most from doing so.

The city fixed effects also imply that α is identified both from variations in density between the cities at origin and destination for movers, and the time changes of density in cities where stayers reside. The worry is then that there can be changes in the city characteristics, which might

[8] For all workers within the same city, city shocks need to be taken into account to estimate standard errors properly (Moulton, 1990).

drive changes in density. For instance, a new road network linking all the main cities may make them more productive and attract more workers. This bias might be of minor importance when estimating from the cross-section as in equation (9) since those new residents attracted by the new road network might represent only a small fraction of city population. When using only the time variation of wages and density, the bias is likely to be much larger since the new road network may explain much of the population growth of cities. Put differently, α as estimated by (10) might be more strongly biased than when using the naive cross-section estimate in (9): the fixed-effect medicine can make the endogeneity disease worse.

To avoid this second problem and lessen the first one, it is possible to impose a city effect for each time period η_{ct} instead of a time-invariant fixed effect η_c. This allows unobserved city characteristics to vary with time. For the effect of density to be identified, we then need to estimate the regression in two stages

$$\log w_{ic(it)t} = \eta_{c(it)t} + u_i + \varepsilon_{ic(it)t}\,, \tag{11}$$

followed by

$$\widehat{\eta}_{ct} = \alpha \log den_{ct} + \varepsilon_{2ct}\,, \tag{12}$$

where the second stage also contain time dummies and the first stage includes time-varying worker characteristics. An alternative to (12) is to average city effects and density across years to estimate $\widehat{\eta}_c = \alpha \log den_c + \varepsilon_{2c}$. Equation (12) (or a close variant) corresponds to the specification adopted by Combes, Duranton, and Gobillon (2008, 2010) or Mion and Nattichioni (2009). Combes, Duranton, and Gobillon (2008) discuss the econometric complications caused by the use of a city fixed effect in the second stage as dependent variable and those created by a large number of fixed effects for both each worker and each city and time period.[9] How does the specification constituted by (11) and (12) fare relative to our three sources of bias?

The endogenous quantity of labour bias is not dealt with (yet). We note that it only matters in the second stage. The endogenous quality of labour bias is addressed, but only partially. Permanent worker characteristics that make them command a higher wage on the labour market are conditioned out. However, the time evolution of these unobserved worker characteristics is ignored. This does not matter if this evolution is idiosyncratic. Nonetheless, changes in worker effects might be caused by where they work and this is not taken into account. We return to this issue very shortly. The shock bias is also partly dealt with. Moving to a given city when expecting a good wage is no longer a source of bias unlike with the naive specification (9). However, moving to a city because a worker receives a 'good offer' when this is correlated with density remains a source of bias. We note however that a 'good offer' here means good conditional on the city and time effect and the worker effect.

According to the results of Combes, Duranton, and Gobillon (2008) or Combes, Duranton, Gobillon, and Roux (2010), the specification made of (11) and (12) yields values of α which are

[9] Note that the estimation procedure in two stages, (11) and (12), has an interesting property: when time-varying worker variables are included in the model, their estimated effect does not depend on the city variables. This is because the city fixed effects are left unspecified in the first stage and this allows for a correlation of unknown form between worker and city variables. We can thus modify the specification of city variables in the second stage without changing anything to the first stage regarding worker variables and their (estimated) effect. This property is not verified when estimating directly variants of equation (10).

about half of those obtained in the naive regression (9). More generally, controlling seriously for individual characteristics using micro-data has large effects on the estimation of agglomeration economies. The sorting of more productive workers into larger cities accounts for a large part of the observed urban wage premium.

An alternative fixed effect strategy would be to adapt Moretti's (2004b) estimation for human capital externalities and impose a fixed effect for each worker and city:

$$\log w_{ic(it)t} = \alpha \log den_{c(it)t} + u_{ic(it)} + \varepsilon_{ic(it)t} \,. \tag{13}$$

With this regression, the effect of density is then estimated for each stayer by looking at how their wage changes when density changes (allowing possibly for experience within a city to be taken into account). Relative to the estimation in equation (11) and (12), there are differences in how this alternative approach deals with the three sources of shocks. The endogenous quantity of labour is ignored, just like above. The difference is that with equation (13) we estimate α from within changes in density and do not rely on cross-city differences in levels. This is especially important because productivity shocks that raise wages are also likely to attract more workers. As for the endogenous quality of labour bias, equation (13) also conditions out time invariant worker effects. It does it through the worker and city effect which is slightly more flexible than a permanent worker effect. Finally, the shock bias is also treated slightly differently and the endogeneity of location is not taken into account.

Returning to the specification made of (11) and (12), a key issue is that the wage growth of workers appears to be stronger in large cities (e.g., Wheeler, 2001). This could be due to faster learning making the worker abilities grow faster in larger cities.[10] To understand how (missing) learning by workers affects fixed-effect estimations, let us consider first a simple example of a worker spending one period in the small city with city effect $\underline{\eta}$, moving for $n+1$ periods to the large city with city effect $\overline{\eta}$, before returning to the small city for one last period. The first period wage in the small city is $\underline{\eta}$. While staying in the large city, the worker learns x after each period. Her wage in the large city is thus $\overline{\eta}$ during her first period, $\overline{\eta} + x$ during her second, and so on until it reaches $\overline{\eta} + n\,x$ during her last period. We consider that what is learnt in the large city can be used in the small city. In this case, the worker's wage is $\underline{\eta} + n\,x$ upon returning to the small city.

For this worker, the estimated effect for the small city is $\underline{\eta} + n\,x/2$ whereas the estimated effect for the large city is $\overline{\eta} + n\,x/2$. The difference between the two is $\overline{\eta} - \underline{\eta}$. Wage differences are thus properly estimated. However wage growth, possibly an important benefit from of cities empirically, is entirely conditioned out. Again, this is not an issue provided we keep in mind that we want to account for wage differences across cities, not measure all the benefits from agglomeration.

Unfortunately, things are likely to be less clear-cut than in this special example. If learning is concave instead of being linear, the estimated effect for the large city will be larger than $\overline{\eta} + n\,x/2$ and the difference in wages between the two cities will thus be overestimated. If the worker spends initially more time in the small city or does not eventually return there, the estimated effect for the

[10] Freedman (2008) considers a first-differenced version (11) where he introduces fixed effects. These fixed effects capture the ability of workers to increase their productivity and possible spatial sorting of 'fast learners'.

small city will be lower and wage differences across the two cities will be overestimated again.[11] It could also be that what is learnt in the large city cannot be entirely transferred back to the small city. Then, it becomes unclear what we mean by wage differences across cities. The non-transferable part of learning in the large city becomes part of its estimated effect, which might be reasonable.

A further source of bias is possible. A worker in the large city may be reluctant to leave it because of learning. Thus, this worker will only take a job in the small city when she receives a very large positive error term (to compensate for foregone learning). The difference in wage when moving to the small city will thus underestimate the true differences between the two cities. This is a specific version of the shock bias.

A final alternative is to opt for a fully structural model that allows for both level and growth effects of cities on worker effects and to attempt to match a number of moments of the data about wage differences both across cities and over time as Baum-Snow and Pavan (2010).

42 *Instruments*

Our main worry at this stage is the endogenous quantity of labour bias. In equation (12) for instance, the estimated city effect $\widehat{\eta}_{ct}$ (which captures wages corrected for worker effects and whatever else is used as controls in the first stage) might be jointly determined with density. A standard solution for that problem is to find one or more variables, refereed to as instruments and noted z_c, that have the following two properties. They can predict density and they are otherwise uncorrelated with the city effects to be explained. Formally (and ignoring time variation), these two conditions are

$$\mathbb{C}\mathrm{ov}(\log den_c, z_c|.) \neq 0, \tag{14}$$

for relevance and

$$\mathbb{C}\mathrm{ov}(\epsilon_{2c}, z_c|.) = 0, \tag{15}$$

for exogeneity when estimating equation (12). The ability of candidate instruments to explain an endogenous variable conditional on the other controls in the regression can be formally assessed. Exogeneity is much harder to establish since it relies on a lack of correlation between the instruments and the unobserved error term. Over-identification tests (also known as tests of over-identifying restrictions) can be conducted when there are more instruments than parameters to estimate. These test implicitly consists in comparing the estimators of the parameters obtained using the different linear combinations of instruments. The hypothesis that the instruments are valid is not rejected when the estimated parameters are close enough from each other. Unfortunately we expect very similar instruments to lead to very similar parameters and thus pass their overidentification tests. This should not be taken as a proof of instrument validity. Overidentification tests are more meaningful when the instruments are very different in nature. In practice, the exogeneity of instruments needs to be carefully discussed by trying to understand what could violate the exclusion restriction (15).

[11] More generally, we note that wage differences across cities are properly estimated because the effects for both the small and large city are equally biased and this disappears when taking the difference between the two. Anything that leads to a difference in bias between the two cities will lead to a biased difference.

A number of instruments have been used in the literature since Ciccone and Hall (1996). Ciccone and Hall's (1996) instruments are mostly long lags of population. The underlying logic is that what drove the location of population two hundred years ago is different from what drives it today (to satisfy the exogeneity condition). At the same time there is much persistence in population patterns, due perhaps to the durability of housing and urban infrastructure (to satisfy the relevance condition). Ciccone and Hall (1996) find only a marginal decline in the elasticity of productivity to density for the US. Replicating their strategy for France and Italy, Combes, Duranton, and Gobillon (2008) and Mion and Nattichioni (2009) find similar results. Following Rosenthal and Strange (2008) in a slightly different context, Combes, Duranton, Gobillon, and Roux (2010) also use the geological characteristics of regions. Fertile soils certainly drove the location of populations when agriculture was a major part of the economy. That geological characteristics still affect the productivity and wages of manufacturing and service sectors is more difficult to imagine. Finally, Combes, Duranton, and Gobillon (2008) also use some measure of geographical periphery like Redding and Venables (2004). The reason for doing this is that more 'central' locations might be more attractive independently from their productivity.[12]

While a reasonable defence of each of these instruments may be provided, none of them is entirely foolproof and could be attacked (and they have). The good news however is that all these instruments appear to give the same answer pointing at a minor endogenous quantity of labour bias.

History, geography and geology are all 'level' instruments for density. If the empirical strategy implies estimating the effects of density 'within' city or time first differencing, the instruments used above become less relevant and their case for exogeneity is much weaker. In one specification, Combes, Duranton, and Gobillon (2008) impose a city fixed effect to equation (12) to estimate $\widehat{\eta}_{ct} = \alpha \log den_{ct} + \eta_c + \varepsilon_{2ct}$. They use lagged city demographics to predict changes in density. These instruments are somewhat weak and it is unclear if initially old cities decline in population because of their age structure or if initially old cities are already poorly productive cities in economic decline. A more promising alternative might be to use the initial sectoral structure of employment interacted with national employment changes in those sectors to predict local evolutions of population. Following Bartik (1991), there is a long tradition of doing this in labour economics to explore a variety of issues. To our knowledge, this approach has not be used to estimate the city size elasticity of wages or productivity.

Often, historical and geological instruments are not available to researchers. Then, they tend to rely on the panel structure of the data to tackle endogeneity issues. Equation (10) can be rewritten in first difference and lagged values of the right-hand side variables may be used as instrument in a GMM approach. This strategy dates back to Henderson (1997) and was later applied, sometimes with some methodological twists, by Henderson (2003), Combes, Magnac, and Robin (2004), and Martin, Mayer, and Meyneris (2008). It is also used to estimate the effect of the market potential on wages in structural models (Mion, 2004; Hanson, 2005).

[12] Ciccone (2002) uses the land area of spatial units to instruments for their density. To the extent that one uses areas that were defined a log time ago, this may be viewed as a proxy for historical populations. Recently defined areas are unlikely to be valid instrument since we expect land area to matter directly in determining wages as argued above.

This approach suffers two caveats. First, as already noted early, identifying agglomeration effects from the time variation of wages and density can be misleading. Second, the validity of instruments is questionable. The variables used as instruments in the GMM approach are all constructed from the lagged values of regressors (such as density). Given the length of those lags, it cannot be argued that all those instruments are exogenous. Meaningful overidentification tests could be conducted if we somehow knew that at least a subset of them is truly exogenous. But it is hard to argue convincingly that this is the case. The fact that overidentification tests are passed may be more indicative of some persistent patterns in the data than having solved endogeneity problems. Put differently, overidentification tests can be passed when the estimators of the parameters obtained using the different linear combinations of instruments all converge to a value which is not the true value of parameters.

43 Modelling location choices explicitly

None of the approaches examined so far deals decisively with the fact that workers choose their location. Typically (and at best) they address the fact that $\mathbb{C}\text{ov}\,(\log den_c, \eta_c) \neq 0$ but not that $\mathbb{C}\text{ov}\left(\log den_{c(i)}, \eta_{c(i)}\right)$ may also be different from zero. Ideally the location choice of workers would need to be modelled and become part of the estimation strategy to deal properly with the endogeneity of the individual location choice in the wage equation.

This is an extremely hard thing to do on two counts. A desirable empirical strategy would first estimate a location choice model followed by a wage equation. The first difficulty is that a choice of location is discrete in nature and this choice is between a large number of locations.[13] Even in a static framework, this raises serious difficulties that can only be overcome through a demanding strategy as demonstrated by Dahl (2002). To our knowledge, there is no methodology able to do this in with panel data.

The second difficulty is perhaps even greater. Proper identification of both a location choice model and a wage equation requires finding a variable that would explain the location without having an effect on the wage. None of the usual determinants of migration is likely to satisfy this exclusion restriction. For instance, married men with children are likely to be less mobile than single childless men. However it is difficult to argue that family status is not associated in any other way with the wage. It is.

Three short-cuts are possible. The first is to note that the location choice equation is non-linear whereas the wage equation is. This implies that one can potentially identify α in a such a system. But this would come entirely from the functional form and is unlikely to be convincing. A more promising alternative is to put more structure behind both the location choice and the determination of wage before structurally estimating such a model like Baum-Snow and Pavan (2010). The quality of the estimation then rest on the quality of the model.[14] The third possibility is to look for particular situations where mobility is artificially restrained or forcibly organised. Au and

[13] Nesting is not a relevant solution. It is hard to argue that migrants choose a city within a class size before choosing a specific one. It is even harder to argue that migrants first choose a region and then a city within it.

[14] See Holmes (2010) for further discussion of structural approaches in agglomeration research.

Henderson (2006) use mobility restrictions in China to assess net agglomeration effects. We discuss this type of research next.

44 *Quasi-experimental evidence*

In areas of economic investigation such as education or welfare policy, experiments are possible and much can be learnt from them. Experimenting with city size or urban density is not feasible in most cases. However, there are particular circumstances that sometimes allow research to replicate experimental conditions through the use of quasi-experiments (or natural experiments).

There is a nascent literature using quasi-experiments to investigate questions related to the estimation of agglomeration economies. Davis and Weinstein (2008) use the exogeneity of the bombing of Japanese cities to search for multiple equilibria in urban structure. Redding and Sturm (2008) exploit the artificial partition of Germany after World War II to assess the importance of market potential. More directly relevant for our purpose, Greenstone, Hornbeck, and Moretti (2010) devise a clever strategy to measure the productivity effects of exogenous labour demand shocks. They use detailed information about bidding wars for large industrial plants in the US. A county that receives one such plant may not be meaningfully compared to the rest of the US in a simple difference-in-difference exercise. However this 'winning' county can be compared to the 'runner-up' county that remained in the bidding process until the very last stage. After providing evidence that winning and runner-up counties are ex-ante observationally very similar, Greenstone, Hornbeck, and Moretti (2010) show that incumbent plants in the winning county experience a large increase in productivity relative to incumbent plants in the runner-up county.[15]

A limitation of this sort of approach is that the 'experiment' is often very limited in scope. Greenstone, Hornbeck, and Moretti (2010) can only assess the productivity effect of a specific shock on other large industrial plants that survive throughout their study period. How much what is learnt from this can be extrapolated to other types of producers and sectors is hard to know. That their results are consistent with other results mentioned above is reassuring. Hopefully future research will be able to provide more evidence of that sort.

45 *Firm selection and firm sorting*

Early on in the derivation of the model of section 2, we made the simplifying assumption that all firms in a city had access to the same technology. We know they do not. Within cities, there are vast differences in measured total factor productivity (Syverson, 2004; Combes, Duranton, Gobillon, Puga, and Roux, 2009). These TFP differences raise two separate issues.

The first is that firms might be immobile but market selection might be at play and this would make the distribution of productivity endogenous. In a nutshell, the argument relies on the idea that denser cities are 'tougher' markets. If market selection eliminates the least productive firms, we should then expect productivity to be on average higher in denser cities. Put differently, market selection might be confounded with agglomeration. Both agglomeration and market selection raise

[15] That productivity remains about stable in the winning county whereas it strongly declines in the runner-up raises some interesting questions.

average productivity in denser cities. To distinguish between market selection and the other forces behind the wage shifter Φ_c, we note that market selection should act by truncating the productivity distribution in denser cities whereas agglomeration forces should have a positive effect on the whole distribution of productivities. Combes, Duranton, Gobillon, Puga, and Roux (2009) develop a new empirical approach to assess greater truncation vs. right-shift or dilation in the distribution of firm productivity for large cities vs. small cities in France. They find no evidence of greater truncation. Future research will confirm or qualify this finding.

Another possibility is that mobile firms might 'sort'. It would be tempting to use a fixed-effect strategy to deal with that issue and attempt to separate firm effects from local effects. However, such approach is likely to be problematic. Only a small proportion of firms move between metropolitan areas every year (Duranton and Puga, 2001). This scarcity of movers makes identification imprecise and suggests that trying to separate firm from local effects might not be meaningful. Furthermore, when firms move, they do not do so randomly. Evidence shows that they move predominantly from larger to smaller cities and, according to theory, might do so following a positive productivity shock. This will bias fixed-effect estimates. Rather than the sorting of firms, it may be more promising to look at the sorting of entrepreneurs (e.g., Michelacci and Silva, 2007).

5. Conclusions

Significant progress has been achieved in the estimation of agglomeration economies over the last 10 years. Standard estimates for the density elasticity of wages now typically range from 0.02 to 0.05. This level of precision may seem satisfactory relative to a number of other important elasticities that the economic profession has attempted to estimate. However, the existence of a consensus is no guarantee of reliability.

We hope further progress will be made on the quality of identification. New plausible instruments are desirable. New quasi-experiments may be waiting to be analysed. Being able to deal seriously with individual location choices is fundamental. The difficulty of this problem suggests that we should remain open-minded in terms of methodology. In particular, structural modelling should be useful in making progress here.

While better estimates are needed, it is also important to explore at greater depth what lies behind the 'average' values that research has uncovered so far. Knowing more about how different types of workers and firms benefit from cities is of first-order importance. There has been work on the sectoral dimension, but heterogeneity by skills and types of firms has only begun to be touched. Hopefully, the dynamics of the agglomeration benefits will also be a vibrant area of progress. Trust in our estimates of agglomeration effects will also be bolstered by knowing more about their exact sources and the channels through which they percolate.

6. References

Alonso, W. (1964). *Location and Land Use; Toward a General Theory of Land Rent.* Cambridge, MA: Harvard University Press.

Au, C.C., and J.V. Henderson (2006). Are Chinese Cities Too Small? *Review of Economic Studies* 73: 549-576.

Bacolod, M., B.S. Blum, and W.C. Strange (2009) Skills in the city. *Journal of Urban Economics* 65: 136-153.

Bartik, T. (1991). *Who Benefits from State and Local Economic Development Policies?* Kalamazoo (MI): W.E. Upjohn Institute for Employment Research

Baum-Snow, N. and R. Pavan (20010). Understanding the city size wage gap. Processed, Brown University.

Behrens, K., G. Duranton, and Frédéric Robert-Nicoud (2010). Productive cities: Sorting, selection and agglomeration. Processed, University of Toronto.

Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review* 46: 213-227.

Ciccone, A., and R.E. Hall (1996). Productivity and the density of economic activities. *American Economic Review* 86: 54-70.

Ciccone, A., and G. Peri (2006). Identifying human-capital externalities: Theory with applications. *Review of Economic Studies* 73: 381-412.

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

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2009). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *CEPR DP* 7191.

Combes, P.-P., G. Duranton, L. Gobillon, and S. Roux (2010). Estimating agglomeration effects with history, geology, and worker fixed-effects. In E.L. Glaeser (ed.) *The Economics of Agglomeration*. Cambridge, MA: National Bureau of Economic Research.

Combes, P.-P., T. Magnac, and J.M. Robin (2004). The dynamics of local employment in France. *Journal of Urban Economics*, 56: 217-243.

Dahl, G.B. (2002). Mobility and the return to education: Testing a Roy model with multiple markets. *Econometrica*, 70: 2367-2420.

Davis, D.R. and D.E. Weinstein (2008). A search for multiple equilibria in urban industrial structure. *Journal of Regional Science*, 48: 29-65.

De La Roca, J. and D. Puga (2010). The dynamic earnings premium of dense cities. Processed, IMDEA and CEMFI (Madrid).

Duranton, G., and D. Puga (2001). Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review* 91: 1454-1477.

Duranton, G., and D. Puga (2004). Micro-foundations of urban agglomeration economies. In Henderson, V., Thisse, J. (Eds.), *Handbook of Urban and Regional Economics*, vol. 4.

Freedman, M.L. (2008). Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry. *Journal of Urban Economics*, 64: 590-600.

Glaeser, E.L. (1999). Learning in cities. *Journal of Urban Economics*, 46: 254-277.

Glaeser, E.L., D.C. Maré (2001). Cities and skills. *Journal of Labor Economics*, 19: 316-342.

Glaeser, E.L., M.G. Resseger (2010). The complementarity between cities and skills. *Journal of Regional Science*, 50: 221-244.

Greenstone M., R. Hornbeck, and E. Moretti (2010). Identifying agglomeration spillovers. *Journal of Political Economy*, 118: forthcoming.

Hanson, G.H. (2005). Market potential, increasing returns, and geographic concentration. *Journal of International Economics*, 67: 1-24.

Henderson, J.V., (1974) The sizes and types of cities. *American Economic Review*, 64: 640-656.

Henderson, J.V. (1988). *Urban Development: Theory, Fact and Illusion*. Oxford University Press.

Henderson, V., A. Kuncoro, and M. Turner (1995). Industrial development in cities. *Journal of Political Economy*, 103: 1067-1090.

Henderson, J.V., (2003). Marshall's scale economies. *Journal of Urban Economics*, 53: 1-28.

Henderson, J.V., (2007). Understanding knowledge spillovers. *Regional Science and Urban Economics*, 37: 497-508.

Henderson, V., (1997). Externalities and industrial development. *Journal of Urban Economics*, 42: 449-470.

Holmes, T.J. (2010). Structural, experimentalist, and descriptive approaches to empirical work in regional economics. *Journal of Regional Science*, 50: 5-22.

Martin, P., T. Mayer, and F. Mayneris (2008). Spatial concentration and firm-level productivity in France. *CEPR DP* 7102.

Melo, P.C., D.J. Graham, and R.B. Noland (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, 39: 332-342.

Michelacci, C. and Silva, O. (2007). Why so many local entrepreneurs? *Review of Economics and Statistics*, 89: 615-633.

Mion, G. (2004). Spatial externalities and empirical analysis: the case of Italy. *Journal of Urban Economics*, 56: 97-118.

Mion, G. and P. Naticchioni (2009). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics*, 42: 28-55.

Mills, E.S. (1967). An aggregative model of resource allocation in a metropolitan area. *American Economic Review (Papers and Proceedings)* 57: 197-210.

Moretti, E. (2004a). Human capital externalities in cities. In Henderson, V., Thisse, J. (Eds.), *Handbook of Urban and Regional Economics*, vol. 4.

Moretti, E. (2004b) Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121: 175-212.

Moulton, B.R. (1990) An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics*, 72: 334-338.

Muth, R.F., (1969). *Cities and Housing.* Chicago: University of Chicago Press.

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science*, 50: 203-219.

Redding, S.J., and D.M. Sturm (2008). The Costs of Remoteness: Evidence from German Division and Reunification. *American Economic Review*, 98: 1766-1797.

Redding, S., and A.J. Venables (2004). Economic geography and international inequality. *Journal of International Economics*, 62: 53-82.

Roback, J. (1982). Wages, rents and the quality of life. *Journal of Political Economy*, 90: 1257-1278.

Rosenthal, S.S., and W.C. Strange (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*, 85: 377-393.

Rosenthal, S.S., and W.C. Strange (2004). Evidence on the nature and sources of agglomeration economies. In Henderson, V., Thisse, J. (Eds.), *Handbook of Urban and Regional Economics*, vol. 4.

Rosenthal, S.S., and W.C. Strange (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64: 373-389.

Rosenthal, S.S., and W.C. Strange (2010). Small establishments/big effects: Agglomeration, industrial organization and entrepreneurship. In E.L. Glaeser (ed.) *The Economics of Agglomeration*. Cambridge, MA: National Bureau of Economic Research.

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

Wheeler, C.H. (2001). Search, sorting, and urban agglomeration. *Journal of Labor Economics* 19: 879-899.