Spatial Sorting:

Why New York, Los Angeles and Detroit attract the greatest minds as well as the unskilled*

Jan Eeckhout[†], Roberto Pinheiro[‡] and Kurt Schmidheiny[§]

September 2010

– preliminary draft –

Abstract

We propose a theory of skill mobility across cities. It predicts the well documented city size–wage premium: the wage distribution in larger and more productive cities first-order stochastically dominates that in less productive cities. Yet, because this premium reflects higher house prices, this does not necessarily imply that this stochastic dominance relation also exists in the distribution of skills. Our model predicts quite to the contrary: instead of first-order, there is second-order stochastic dominance in the skill distribution. The demand for skills is non-monotonic as our model predicts a "Sinatra" as well as an "Eminem" effect: both the very high and the very low skilled disproportionately sort into the biggest cities, while those with medium skill levels sort into small cities. Based on our theory, the pattern of spatial sorting is explained by a simple technology with varying elasticity of substitution by skill. Using CPS data on wages and Census data on house prices, this technology with the elasticity of substitution decreasing in skill density is consistent with the observed patterns of skills.

Keywords. Matching theory. Sorting. General equilibrium. Population dynamics. Cities. Wage distribution.

*We are grateful to numerous colleagues for insightful comments. We also benefitted from the feedback of several seminar audiences. Eeckhout gratefully acknowledges support by the ERC, Grant 208068.

[†]ICREA-UPF Barcelona and UPenn jan.eeckhout@upf.edu.

[‡]University of Colorado, Roberto.Pinheiro@colorado.edu.

[§]Department of Economics, Universitat Pompeu Fabra, kurt.schmidheiny@upf.edu.

"*If I can make it there I'll make it anywhere...*" (Frank Sinatra – New York, New York)

"*Rock Bottom, yeah I see you, all my Detroit people*" (Eminem – Welcome 2 Detroit)

1 Introduction

New York, NY. Making it there rather than in Akron, OH is the ultimate aim of many professionals. And this is true for many trades and skills: artists, musicians, advertising and media professional, consultants, lawyers, financiers,... While there are certainly notable exceptions (the IT sector comes to mind), most people can provide casual evidence that the skill level in the top percentiles of NY and large cities in general is higher than anywhere else. Yet, to date there is little or no empirical evidence to back this up. While there is certainly ample evidence of a city-size wage premium, there is little evidence of sorting of both the skilled and the unskilled across different size cities (see for example Baum-Snow and Pavan (2010a,b)).

In this paper we show that there is indeed evidence that disproportionately more skilled citizens locate in larger cities. However, we provide a key new insight: larger cities also disproportionately attract lower skilled agents. And it turns out that large cities like New York and Detroit in that respect are more similar to each other than to small cities. For example, we find that New York city attracts the low skilled as well as the high skilled. Most who do not live there are usually most familiar with Manhattan, but there is a huge low skill contingent in the Bronx and Queens. And likewise Detroit, the 14th largest metro area in 2009 has disproportionately many low skilled individuals and a reputation for inner city poverty and low skilled work. Yet, quite contrary to most common preconceived notions, the Detroit metro area also disproportionately attracts high skilled individuals. We show that there is a systematic pattern of fat tails in the skill distribution of large cities. To our knowledge, the relation between city size and the disproportionate presence of both the high and the low skilled has not been documented in the literature.

As will become apparent, it is precisely the fact that large cities attract the low skilled citizens as well as the high skilled that explains why to date little or no evidence has been found on sorting. Our finding thus also encompasses a methodological contribution. In the literature, skills are often partitioned into two classes,[1] which allows for inference of a linear approximation when the underlying relation is monotonic. However, in the case of skills across cities, the equilibrium skill quantity relative to the economy-wide quantity is essentially non-monotonic: the skill demand in large cities relative to small ones is U-shaped as large cities disproportionately attract both high and low skilled workers, but not those with medium skill levels. There is no hope to satisfactorily identify this non-monotonic relation with two points only. Our approach is to allow for many skill classes, in fact, as many as there are observations. We can thus characterize a smooth distribution of skills.

[1] The same is true when the focus is on occupations instead of skills. Gould (2007) partitions the set of workers into blue and white collar occupations.

The city-size wage premium is well documented. For example, the gap between average wages in the smallest cities in our sample (with a population around 160,000, more than 100 times smaller than New York) and the largest cities is 25%. Below in Figure 1.A, we plot a kernel of the wage distribution of those living in all cities larger than 2.5 million inhabitants and that of those in cities smaller than one million inhabitants. Not only are average wages higher, there is a clear first-order stochastic dominance relation. At all wage levels, more people earn less in small cities than in large cities. This clearly indicates that there is a city-size wage premium across the board. Of course, larger cities tend to be more expensive to live, so in order to be able to compare skill distributions, we need to adjust for house prices. Identical agents will make a location choice based on the utility obtained, which depends both on wages and the cost of housing. Indifference for identical agents will therefore require equalizing differences. We use homothetic preferences to adjust for housing consumption and construct a house price index based on a hedonic regression to calculate the difference in housing values across cities. The resulting distribution of utilities will therefore be isomorphic to the distribution of skills in a world with full mobility and no market frictions. Figure 1.B displays the kernel of the skill distribution. The skill distribution in larger cities has fatter tails both at the top and at the bottom of the distribution. Large cities disproportionately attract more skilled and more unskilled workers. Figure 1.D illustrates that this implies a non-monotonic relative demand for skills in equilibrium, whereas the relative wages are monotonic (Figure 1.C).

We propose a simple theory of city choice and heterogeneous skills. The role of citizens is to earn a living based on a competitive wage, and under perfect mobility, agents need to be indifferent between different consumption-housing bundles, and therefore between different wage-house price pairs. Wages are generated by firms that compete for labor and that have access to a city-specific technology summarized by that city's total factor productivity (TFP). Output is produced with heterogeneous labor inputs. It values higher skills, but the marginal product of labor is decreasing in the number of workers hired of that skill level. As a result, given a vector of wages, firms want to hire a combination of different skills. Apart from differences in TFP, the technology is the same across cities. The labor aggregating technology of different skilled workers is the Varying Elasticity of Substitution (VES) technology. It is similar to the Constant Elasticity of Substitution (CES) technology, but the exponent on the quantity (and therefore the marginal product) is allowed to vary by skill. As a result, the elasticity of substitution is varying.

For the general technology – both CES and VES –, the size of the city will typically be increasing with TFP and we can establish that wages are higher in larger cities. Firms in high TFP cities are more productive and can attract workers paying higher wages. In equilibrium, labor demand will also push up house prices. The cititzens' location decision will equalize utility and a worker of a given skill will be indifferent between a high wage, high house price city and a low wage, low house price city.

The shape of the distribution is crucially determined by the technology. In the benchmark case of CES, cities of different TFP have different population sizes, but the distribution of skills is the

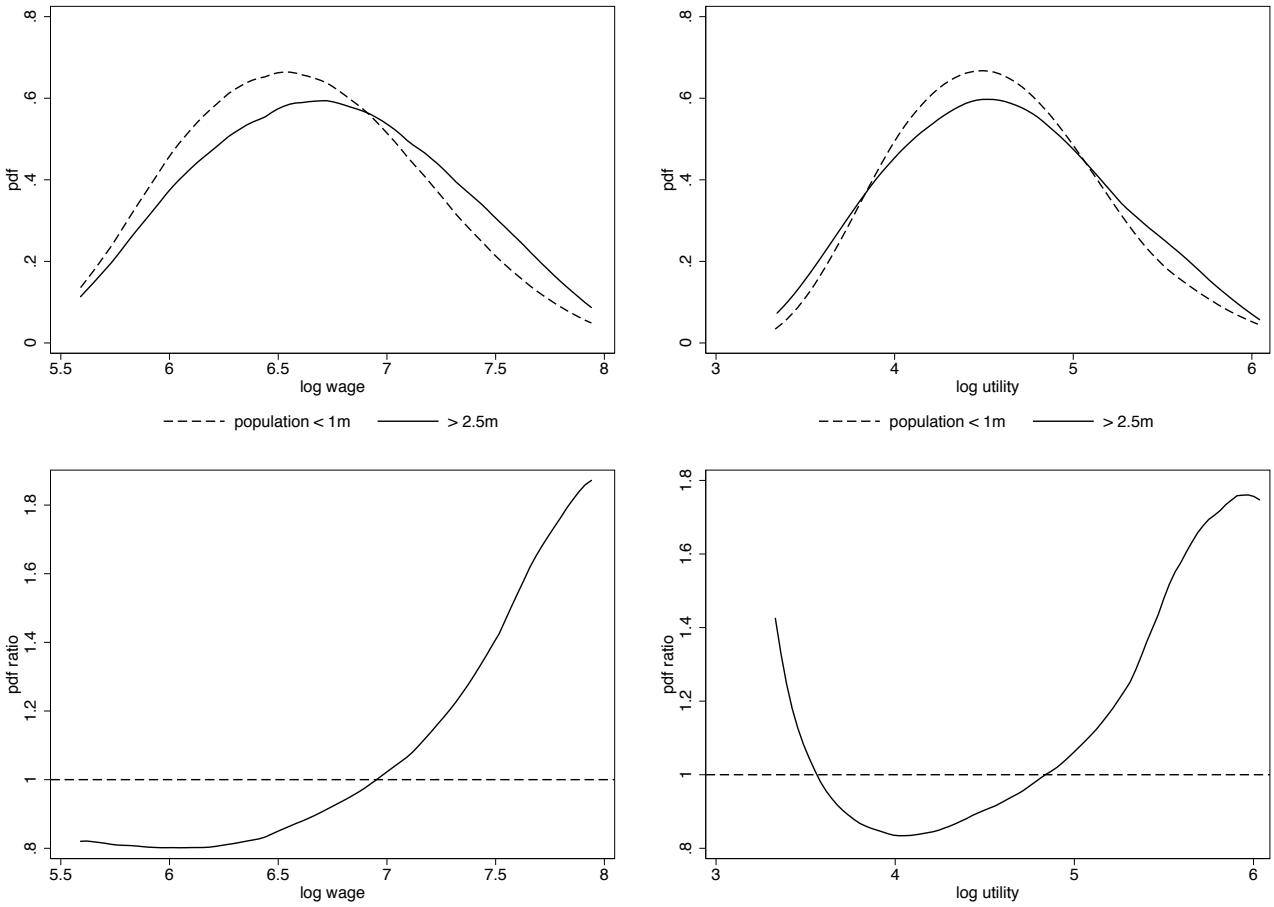

Figure 1: Left-to-right-top-to-bottom. A. Wage distribution for small and large cities; B. Skill distribution for small and large cities; C. Relative wages between large and small cities; D. Relative demand for skills between large and small cities.

same across cities, and for that matter across the entire economy. In contrast, when the elasticity of substitution is decreasing in the measure of a given skill, then larger cities have skill distributions with fatter tails.

The primitive of our model is city-specific total factor productivity. Like in models of growth, this is likely to be generated in part endogenously. It thus consists of both a pure productivity portion as well as a portion that is ascribed to externalities from agglomeration, the presence of amenities,...[2] We propose a micro-foundation for this particular VES technology. Consider the standard CES technology with in addition knowledge spillovers between workers of different skill levels. People randomly meet someone with a different skill level in the city and benefit from some newly acquired knowledge as a result. They can only learn from differently skilled types. As a result, those in scarce skills are

[2]While realistically it is determined endogenously, for the purpose of our model we take it is as given and assume it is not affected directly by investment by individuals or institutions. For example, a local government may be able to affect its city-specific TFP through investment, for example in local transportation or the construction of an airport.

more likely to meet a different skilled type and therefore benefit more from the spillover than those in abundant skills. As a result, the marginal product of the scarce skilled types is larger.

The key feature of our approach is to use revealed preference choice of location and wages paid to back out skills. This is in contrast to the commonly used approach of using observable skills (levels of education, test scores,...). One reason for doing things this way is that observable skills only explain a fraction of skills (see for example Keane and Wolpin (1997)). Moreover, skill categories are typically very coarse, and given the non-monotonicity observed using our approach – that relative to small cities, the equilibrium demand for skills in large cities is U-shaped in skill – identifying this non-monotonicity from coarse data is challenging for the obvious reasons mentioned above.

In addition to our wage based measure, we also derive the skill distribution based on observable skill, either by schooling category or actual years of schooling, derived from self-reported educational attainment in the CPS data. We find the same qualitative prediction of fatter tails and second order stochastic dominance in larger cities. With the objective of providing external validation, we further use the observable skill categories to calculate the determinants of the technology and how they vary by skill. We find that the marginal product (and therefore the elasticity of substitution) is higher for the scarce skills, i.e., both for the highest and the lowest skills. This U-shaped pattern is precisely the driving force in our theory behind the fat tailed skill distribution.

2 Related Empirical Literature

The model we propose builds on the urban location model in Eeckhout (2004) and Davis and Ortalo-Magné (2009) where identical citizens who have preferences over consumption and housing choose a city in order to maximize utility. Because of differences in productivity across cities, wages differ and house prices adjust in function of the population size of the city. Productivity differences are due to TFP and agglomeration effects. Given perfect mobility and identical agents, utility equalizes across cities. Here we add heterogeneity in the inputs of production (skills) which gives rise to a distribution of skills within the city. The production technology aggregates different skilled inputs within a firm without assuming a constant elasticity of substitution technology as in Eeckhout and Pinheiro (2010). Equilibrium is determined by the sorting decision of agents.

Our empirical exercise compares the entire distribution of real wages across cities of different sizes in the United States. We are not the first to study wages across cities. Behrens, Duranton and Robert-Nicoud (2010) regress log nominal wages on log city size across 276 MSA areas using 2000 Census data. They find an average urban premium of 8% without controlling for talent, measured by education, and 5% when controlling for it. In addition, they regress housing costs on city size using both rental prices and an index formed of rental price and housing values of owner-occupied units. They find similar coefficients for housing costs as for nominal wages, suggesting that there is no substantial difference in real wages. This is consistent with our finding that the mean of house-price adjusted wages is the same.

They do not analyze the higher variance in larger cities.

Moretti (2010) calculates real urban wages using wage data across 315 MSAs using the 1980, 1990 and 2000 Census and focusses on the change in inequality over time. He compares two different local price indices, one based on rental prices only and one including local differences for consumption goods. The latter being based on BLS data for the 23 largest MSAs. Real wages are calculated as the nominal wage divided by one of local price indices and then used to estimate the wage difference between workers with a high school degree and workers with college or more. He finds that the cost of living has gone up more for college graduates and as a result, real wage inequality has increases less than nominal wage inequality.

Baum-Snow and Pavan (2010b) use wages from the 1980, 1990 and 2000 Census (5% PUMS) plus the 2005-2007 American Community Surveys (ACS). Wages are deflated for inflation but not for local differences in housing prices. They document that the variance of hourly wages in increasing with city size. This relationship is weak in 1980 and increases steadily until 2007. Baum-Snow and Pavan (2010a) in addition use the local cost-of-living index from the Commerce Research Association (ACCRA) for 244 metropolitan areas and 179 rural counties 2000 to 2002. Real wages are then used to estimate both log wage level and log wage growth differences across city size groups. They also control for different education groups. They find that real wages are up to 30% higher in MSAs of over 1.5 million inhabitants than rural areas. Baum-Snow and Pavan (2010a) does not study the variance of real wages.

Albouy (2008) calculates real urban wages for 290 MSAs using the 2000 Census (5% IPUMS). Nominal wages are deflated using rental prices from the Census and local prices for consumption goods. The ACCRA Cost-of-Living index is the basis of the latter but not directly used because of its limited quality. Albouy regresses the ACCRA index on local rental prices and uses the predictions as index for local cost-of-living differences. Differences in real wages across MSAs are interpreted as quality-of-live differences. He finds that controlling for local differences in federal taxes, non-labor income and observable amenities such as seasons, sunshine, and coastal location, quality of life does not depend on city size.

All this body is consistent with our finding that the average of the skill distribution is remarkably constant across different size cities. Of course, that does not allow us to conclude that there is no sorting or that there is sorting of high skilled workers in large cities and low skilled workers in small cities. As we will show below, quite to the contrary. The mean is constant across cities of different size, but the variance is significantly increasing. The latter indicates an important role of sorting of high and low types into large cities and of medium types into small cities.

3 The Model

Population. Consider an economy with heterogeneously skilled workers. Workers are indexed by a skill type i. For now, let the types be discrete: $y_i, i \in \mathcal{I} = \{1, ..., I\}$, where the skill level y_i is increasing

in i. Denote the country-wide measure of skills of type i by M_i. Let there be J locations (cities) $j \in \mathcal{J} = \{1, ..., J\}$. The amount of land in a city is fixed and denoted by H_j. Land is a scarce resource, and it will be assumed that the total stock of land available is for residential use.

Preferences. Citizens of skill type i who live in city j have preferences over consumption c_{ij}, and the amount of land (or housing) h_{ij}. The consumption good is a numeraire good with price equal to one. The price per unit of land is denoted by p_j. We think of the expenditure on housing as the flow value that compensates for the depreciation, interest on capital,... In a competitive rental market, the flow payment will equal the rental price.[3] A worker of type i has consumer preferences in city j that are represented by:

$$u(c_{ij}, h_{ij}) = c_{ij}^{1-\alpha} h_{ij}^{\alpha}$$

where $\alpha \in [0, 1]$. Workers and firms are perfectly mobile, so they can relocate instantaneously and at no cost to another city. Because workers with the same skill are identical, in equilibrium each of them should obtain the same utility level wherever they choose to locate. Therefore for any two cities j, j' it must be the case that:

$$u(c_{ij}, h_{ij}) = u(c_{ij'}, h_{ij'}),$$

for all skill types $\forall i \in \{1, ..., I\}$.

Technology. Cities differ in their total factor productivity (TFP) which is denoted by A_j. We treat this as exogenous and represents a city's productive amenities, infrastructure, historical industries,...[4]. In each city, firms compete to operate in this market. Firms all assumed to be identical and to have access to the same, city-specific TFP. Output is produced from choosing the right mix of different skilled workers i. For each skill i, a firm in city j chooses a level of employment m_{ij} and produces output

$$A_j \sum_{i=1}^{I} (m_{ij})^{\gamma_i} \, y_i^{\beta},$$

where γ_i is skill-dependent. When γ_i is constant for all i, this technology is the standard CES (constant elasticity of substitution). Because γ_i is skill-varying, we refer to this technology as VES (varying elasticity of substitution). Firms pay wages w_{ij} for workers of type i. It is important to note that wages will depend on the city j because citizens freely locate between cities not based on the highest wage, but given house price differences, based on the highest utility.

Entry into the market entails a cost kp_j. We assume that the entry cost are city specific and depend on housing prices. Firms need to rent housing space for production and house prices affect the

[3] We will abstract from the housing production technology, for example we can assume that the entire housing stock is held by a zero measure of landlords.

[4] We assume this exogenous because our focus is on the allocation of skills across cities, but one can easily think of this being the outcome of investment choices made by firms, local governments,...

expenditure (e.g., municipal taxes) that a firm incurs to finance infrastructure. Competitive entry will drive down profits to zero, which are given by:

$$A_j \sum_{i=1}^{I} (m_{ij})^{\gamma_i} y_i^{\beta} - \sum_{i=1}^{I} w_{ij} m_{ij} - k p_j = 0.$$

The measure of firms entering the market in city j is denoted by N_j and is determined by this zero profit condition and the market clearing conditions below.

Market Clearing. In the housing market of each city j, market clearing in the housing market requires:

$$\sum_{i=1}^{I} h_{ij} m_{ij} = \frac{H_j}{N_j}, \forall j.$$

In the country-wide market for skilled labor, markets for skills clear market by market:

$$\sum_{j=1}^{J} N_j m_{ij} = M_i, \ \forall i.$$

4 The Equilibrium Allocation

The Citizen's Problem. Within a given city j and given a wage schedule w_{ij}, a citizen chooses consumption bundles $\{c_{ij}, h_{ij}\}$ to maximize utility subject to the budget constraint (where the tradable consumption good is the numeraire, i.e. with price unity)

$$\begin{aligned} \max_{\{c_{ij}, h_{ij}\}} u(c_{ij}, h_{ij}) &= c_{ij}^{1-\alpha} h_{ij}^{\alpha} \\ \text{s.t. } c_{ij} + p_j h_{ij} &\leq w_{ij} \end{aligned}$$

for all i, j. Solving for the competitive equilibrium allocation for this problem we obtain:

$$\begin{aligned} c_{ij} &= (1-\alpha) w_{ij} \\ h_{ij} &= \alpha \frac{w_{ij}}{p_j} \end{aligned}$$

Substituting the equilibrium values in the utility function, we can write the indirect utility for a type i as:

$$U_i = \alpha^{\alpha} (1-\alpha)^{1-\alpha} \frac{w_{ij}}{p_j^{\alpha}} \Rightarrow w_{ij} = U_i p_j^{\alpha} K,$$

where $K = 1/\alpha^{\alpha} (1-\alpha)^{1-\alpha}$. This allows us to link the wage distribution across different cities j, j'. Wages across cities relate as:

$$w_{ij} = w_{ik} \left(\frac{p_j}{p_{j'}} \right)^{\alpha}.$$

The Firm's Problem. Given the city production technology, a firm's problem is given by:

$$\max_{m_{ij},\forall i} A_j \sum_{i=1}^{I} (m_{ij})^{\gamma_i} y_i^{\beta} - \sum_{i=1}^{I} w_{ij} m_{ij} - kp_j$$
$$\text{s.t. } m_{ij} \geq 0, \forall i$$

The first-order condition is:[5]

$$\gamma_i A_j (m_{ij})^{\gamma_i - 1} y_i^{\beta} = w_{ij}, \forall i.$$

All firms are price-takers and do not affect wages. Wages are determined simultaneously in each submarket i, j. Even without fully solving the system of equations for the equilibrium wages, observation of the first-order condition reveals that productivity between different skills i in a given city are governed by two components: 1. the productivity y_i of the skilled labor and how fast it changes between different i (determined by β); and 2. the measure of skills m_{ij} employed (wages decrease in the measure employed from the concavity of the technology).

We take the view that wages are monotonic in skills.[6] Since utility is increasing in skills and equalized across cities, the utility distribution is therefore a monotonic transformation of the skill distribution. The skill distribution may have a different shape than the utility distribution, but its ordinal features are preserved. In particular, if we compare two utility distributions, the densities of which intersect twice, then also the skill densities will intersect twice. In other words, if there are fat tails in the utility distribution, then there are also fat tails in the skill distribution.

In order to simplify the exposition and the derivations, we now proceed the analysis with two cities $j = 1, 2$ and any number I of skills. From the labor market clearing condition and using the first-order condition in both cities we can substitute for m_{ij} to obtain:

$$\left(\frac{w_{ij}}{\gamma_i A_j y_i^{\beta}} \right)^{\frac{1}{\gamma_i - 1}} = \frac{M_i}{N_j} - \frac{N_{j'}}{N_j} \left(\frac{w_{ij'}}{\gamma_i A_{j'} y_i^{\beta}} \right)^{\frac{1}{\gamma_i - 1}}.$$

[5] In what follows, the non-negativity constraints on m_{ij} will be dropped since the marginal product at zero tends to infinity whenever γ_i and A_j are positive.

[6] Whenever the density is upward sloping, then the second effect will imply that higher skilled workers will see an increase in productivity only provided the first effect (productivity) is large enough. In fact, when $\beta = 0$, the effect of skills is only through m_{ij} and wages will be decreasing if the density is increasing. Likewise, the first effect completely dominates when β tends to infinity. For every economy, there exists a critical $\beta^{\star}$ such that productivity (and therefore equilibrium wages) is increasing in skill i in every city j. We will in what follows therefore assume that productivity is monotone in skills: $\beta > \beta^{\star}$. This is without loss of generality if one interprets the order of skills to be determined by the marginal product. Suppose $\beta < \beta^{\star}$ and for some i wages decrease in skill i, then one can suitably reorder skills such that wages that pay lower wages are assigned a lower index such that under the new index wages are increasing. This might be the case for the average artist or architect for example, who in terms of years of schooling are more skilled than accountants, yet they earn less (either because they are abundant or because the productivity is low). In our view of skills, the account would be more skilled than the artist.

Then we can write the wage ratio as:

$$\frac{w_{i1}}{w_{i2}} = \left\{ \frac{\frac{A_1}{A_2}\left(\frac{w_{i1}}{\gamma_i A_1 y_i^{\beta}}\right)^{\frac{1}{\gamma_i - 1}}}{\frac{M_i}{N_2} - \frac{N_1}{N_2}\left(\frac{w_{i1}}{\gamma_i A_1 y_i^{\beta}}\right)^{\frac{1}{\gamma_i - 1}}} \right\}^{\gamma_i - 1}$$

Since from perfect mobility in the consumer's optimization problem we have utility equalization and as a result $\frac{w_{i1}}{w_{i2}} = \left(\frac{p_1}{p_2}\right)^{\alpha}$, we can write the equilibrium employment levels for each skill i in both cities as:

$$m_{i1} = \frac{\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i - 1}}}{1 + \frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i - 1}}} \frac{M_i}{N_2}$$

$$m_{i2} = \frac{1}{1 + \frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i - 1}}} \frac{M_i}{N_2}$$

We can then express the equilibrium wages explicitly as

$$w_{i1} = \left\{ \frac{\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i - 1}} \frac{M_i}{N_2}}{1 + \frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i - 1}}} \right\}^{\gamma_i - 1} \gamma_i A_1 y_i^{\beta}$$

and similarly for w_{i2}.

This wage equation allows us to establish the relation between skills and wages in the data. It also confirms that there is a one-to-one relation between skills and equilibrium utility U_i. Subject to a transformation between w_{ij} and y_i we can therefore without loss of generality express skill qualitatively by $U_i = \alpha^{\alpha}(1-\alpha)^{1-\alpha}\frac{w_{ij}}{p_j^{\alpha}}$.

Finally, the equilibrium is fully specified once we satisfy market clearing in the housing market in each city and we pin down the measure of firms N_j from the zero profit condition. Expressing the housing market clearing condition as a ratio, we get together with the zero profit conditions:

$$\frac{\sum_{i=1}^{I}\left(\frac{M_i}{N_2}\right)^{\gamma_i}\gamma_i y_i^{\beta}\left\{\frac{\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}}}{1+\frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}}}\right\}^{\gamma_i}}{\sum_{i=1}^{I}\left(\frac{M_i}{N_2}\right)^{\gamma_i}\gamma_i y_i^{\beta}\left\{\frac{1}{1+\frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}}}\right\}^{\gamma_i}} = \frac{H_1}{H_2}\frac{A_2}{A_1}\frac{N_2}{N_1}\frac{p_1}{p_2} \tag{1}$$

$$\sum_{i=1}^{I}\left(\frac{M_i}{N_2}\right)^{\gamma_i}(1-\gamma_i)\,y_i^{\beta}\left(\frac{\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}}}{1+\frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}}}\right)^{\gamma_i} = \frac{k}{A_1}p_1 \tag{2}$$

$$\sum_{i=1}^{I}\left(\frac{M_i}{N_2}\right)^{\gamma_i}(1-\gamma_i)\,y_i^{\beta}\left(\frac{1}{1+\frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}}}\right)^{\gamma_i} = \frac{k}{A_2}p_2 \tag{3}$$

In what follows we refer to these as the three equilibrium conditions (1)–(3).

The Main Theoretical Results. First we consider the benchmark case where the technology has a constant elasticity of substitution (CES). This provides a benchmark for our main findings about the distribution of skills across cities.

Theorem 1 *CES technology. If $\gamma_i = \gamma$ for all i, then the skill distribution across cities is identical.*

Proof. In Appendix. ■

The CES technology implies that cities have identical skill compositions. This is due to the homotheticity of the CES technology: the marginal rate of technical substitution is proportional to total employment, and as a result, firms in different cities and with different technologies will employ different skills in the same proportions.

We now establish the relation between TFP and city size. Denote by S_j the size of city j where $S_j = \sum_{i=1}^{I} N_j m_{ij}$. When cities have the same amount of land, we can establish the following result for a general technology.

Proposition 1 *City Size and TFP. Let $A_1 > A_2$ and $H_1 = H_2$, then $S_1 > S_2$.*

Proof. In Appendix. ■

We establish this result for cities with identical supply of land. Clearly, the supply of land is important in our model since a city with an extremely tiny geographical area would drive up housing prices all else equal. This may therefore make it more expensive to live even if the productivity is lower. Because in our empirical application we consider large metropolitan areas (NY city for example

includes large parts of New Jersey and Connecticut with relatively low population density), we believe that this assumption is without much loss of generality.[7]

We now proceed to showing the main result. We already know that more productive cities are larger, but that does not necessarily mean that the distribution of skills in larger cities differs from that in smaller cities. In fact, it depends on the technology. We know from Theorem 1 that for the CES technology the large cities have exactly the same distribution as the smaller cities.

We therefore make the following assumption on how the coefficient γ_i varies with i. Below, we provide a simple micro-foundation for this assumption.

Assumption 1 *γ_i is decreasing in the economy-wide density of skill i.*

In other words, scarce skills have a higher γ_i than abundant skills. This is illustrated in Figure 4. It is important to note here that γ_i does not depend on the firm's employment in skill i. This would affect the firm's first order condition as it will take into account how the marginal product is affected by the change in m_{ij}. Because the firm is infinitesimally small relative to the market, it takes the aggregate employment level as given.

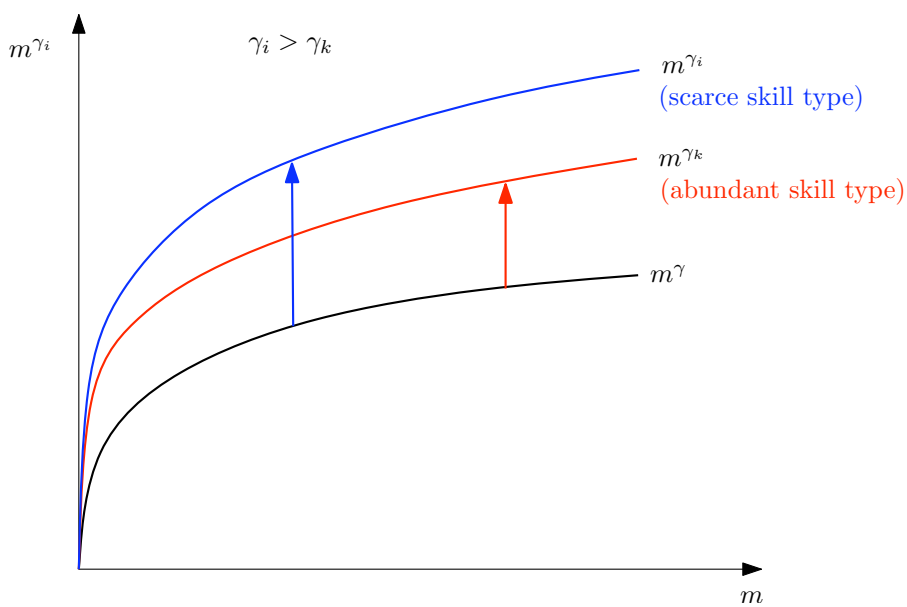

Figure 2: Scarce skill types have a higher γ_i, implying a higher level of productivity *and* a higher a marginal productivity.

We can now establish the main theorem characterizing the skill distribution across firms:

[7] In fact, the equal supply of housing condition is only sufficient for the proof, not necessary. However, our model does not speak to the important issue of within-city geographical heterogeneity, as analyzed for example in Lucas and Rossi-Hansberg (2002). In our application, all heterogeneity is absorbed in the pricing index by means of the hedonic regression.

Theorem 2 *Fat Tails. Consider a symmetric, uni-modal skill distribution economy-wide. Then under Assumption 1, $A_1 > A_2$, and $H_1 = H_2$, the skill distribution in larger cities has fatter tails.*

Proof. In Appendix. ■

To see the nature of this result, consider the benchmark of CES. Homotheticity implies that even though the level of employment differs across skills, firms will always choose to hire different skills in exactly the same proportions for a given wage ratio. Since house prices affect all skills within a city in the same way, the wage ratio is unaffected. Instead, when the marginal product is higher for low and high skilled workers relative to the medium skilled, the higher TFP cities have a comparative advantage in hiring those scarce skills relative to the abundant skills. Due to the complementarity between TFP and the labor aggregator, firms in large cities have a comparative advantage in hiring scarce skills where the marginal product is highest.

We now discuss some further implications of the model.

Housing Consumption and Expenditure. It is immediate from our model that in large cities, citizens will spend more on housing, yet they will consume less of it.

Proposition 2 *Let $A_1 > A_2$ and $H_1 = H_2$. For a given skill i, expenditure on housing $p_j h_{ij}^\star$ is higher in larger cities. The size of houses $h_{ij}^\star$ in larger cities is smaller.*

Proof. From the consumer's problem, we have: $p_j h_{ij} = \alpha w_{ij}$. Then, since we established in the proof of Proposition 1, that $w_{i1} > w_{i2}$, we must have $p_1 h_{i1} > p_2 h_{i2}$, $\forall i$. Similarly, from the same equality in the consumer's problem, we have $h_{ij} = \alpha w_{ij}/p_j$. Again, from the proof of Proposition 1, we have:

$$\frac{w_{i1}}{p_1} < \frac{w_{i2}}{p_2}$$

which implies $h_{i1} < h_{i2}$. ■

Then given homothetic preferences for consumption, it immediately follows that:

Corollary 1 *Expenditure on the consumption good is higher in larger cities.*

Our model predicts that expenditure on both housing and consumption is higher in larger cities, though the equilibrium quantity of housing $h_{ij}^\star$ is lower. As cities become larger (or as the difference in TFP increases), at all skill levels total income increases and therefore total expenditure increases. Because house prices increase as well, there will be substitution away from housing to the consumption good. As a result, inequality in consumption expenditure will increase.

Firm Size. Our model is ambiguous when it comes to predicting firm size across cities. By assumption, there is a representative firm within a given city, and the firm size in city j is given by $\sum_i m_{ij}$. Due to the free entry condition for firms and the ensuing general equilibrium effects, the firm size can

Proposition 3 *Firm size is ambiguous across different cities.*

Proof. We know that $S_j = N_j \sum_i m_{ij}$. Therefore we can write relative firm size as

$$\frac{\sum_i m_{i1}}{\sum_i m_{i2}} = \frac{S_1}{S_2}\frac{N_2}{N_1}.$$

The ratio of populations $S_1/S_2 > 1$. We also know that

$$\frac{N_1}{N_2} < \frac{A_2}{A_1}\frac{p_1}{p_2},$$

however the RHS can be either smaller or larger than 1 (the TFP ratio is smaller and the price ratio is larger than one). It follows that

$$\frac{\sum_i m_{i1}}{\sum_i m_{i2}} \lesseqgtr 1.$$

■

In the data section, we will verify how firm size changes across cities.

Labor Productivity and TFP. Even though large cities attract low skilled workers, those low skilled workers are more productive in large cites. In fact, as we pointed out earlier, the wage and therefore labor productivity in the largest cities is on average 25% larger than that in the smallest cities in our sample. Even under CES, more low skilled workers go to large cities because their productivity is higher there (though they do this to proportional the high skilled under CES). When the elasticity is varying, then in addition, the larger marginal product γ_i for scarce skills relative to abundant skills makes the labor productivity of the scarce skills even larger in large cities.

Given the wage distribution within the city, house prices and the city size, we can infer information about the underlying productivity. For example, for the two-city economy, all else equal, an increase in the city size of the largest city is driven by an increase in TFP in that city. Our model is static, and therefore silent on the evolution of wages across cities. Nonetheless, based on these comparative statics, one can infer from the evolution of wages and skills across cities how productivity has evolved over time across cities.

5 The Empirical Evidence of Fat Tails

5.1 Empirical Strategy

We use the one-to-one relation between skills and equilibrium utility to back out the skill distribution from easily observable variables. The worker's indirect utility in equilibrium is independent of the city,

given perfect mobility, and assuming Cobb-Douglas preferences, it satisfies

$$U_i = \alpha^{\alpha} (1-\alpha)^{1-\alpha} \frac{w_{ij}}{p_j^{\alpha}} \tag{4}$$

where we need to observe the distribution of wages w_{ij} by city j, the housing price level p_j by city and the budget share of housing α.

5.2 Data

The analysis is performed at the city level. We define a city as a Core Based Statistical Area (CBSA), the most comprehensive functional definition of metropolitan areas published by the Office of Management and Budget (OMB) in 2000. See Table 1 for examples of cities and their 2009 population.

[Insert Table 1 here]

We use wage data from Current Population Survey (CPS) for the year 2009. We observe weekly earnings for 102,577 full-time workers in 257 U.S. metropolitan areas. CPS wages are top-coded at around $150,000 which we will take into account in the statistical analysis.

Local housing price levels are estimated using the 5% Public Use Microsample (PUMS) of the 2000 U.S. Census of Housing. We observe monthly rents for 3,274,198 housing units and assessed housing values for 7,680,728 owner-occupied units in 533 CBSAs. The Census also reports the number of rooms and bedrooms, the age of the structure, the number of units in the structure and whether the unit has kitchen facilities. City specific price indices from the Federal Housing Finance Agency (FHFA) based on the Case and Shiller (1987) repeat sales method are used to adjust for 2000-2009 growth in housing prices.

See the data appendix for more details on data source, sample restrictions and variables.

5.3 Wage distribution

Figure 3 shows the distribution of weekly wages for full-time earners both in cities with a population of more than 2.5 million and cities with population between 100,000 and 1 million. We clearly see that wages in larger cities are higher and that the top tail of the distribution is substantially bigger in large cities. A simple t-test shows that wages in large cities are 13.4% higher than in small ones ($t = 28.7, p < 0.000$). Controlling for right censoring from top-coding and weights in a censored (tobit) regression leads to almost exactly the same comparison: Δ log wage= 13.4% (robust $t = 25.54$, $p < 0.000$).

The above partitioning of wages into a group of small cities and a group of large cities ignores substantial differences across different cities of similar size. We therefore also relate the wage distribution of individual cities to city size. We estimate the mean and the standard deviation of the right-censored

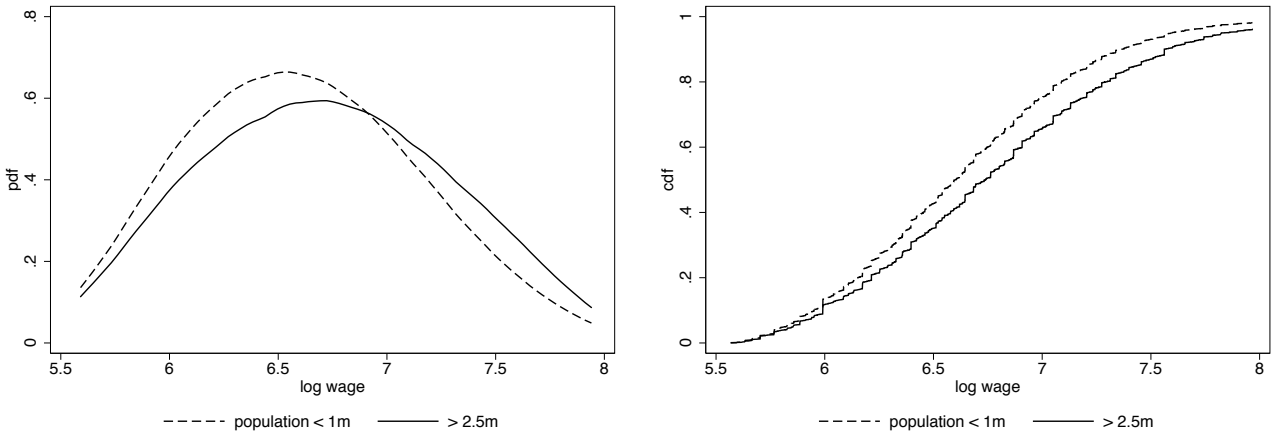

Figure 3: Wage distribution for small and large cities. Full-time wage earners from 2009 CPS. A. Kernel density estimtates (Epanechnikov kernel, bandwidth = 0.2), adjusted for top-coding; B. Empirical CDF.

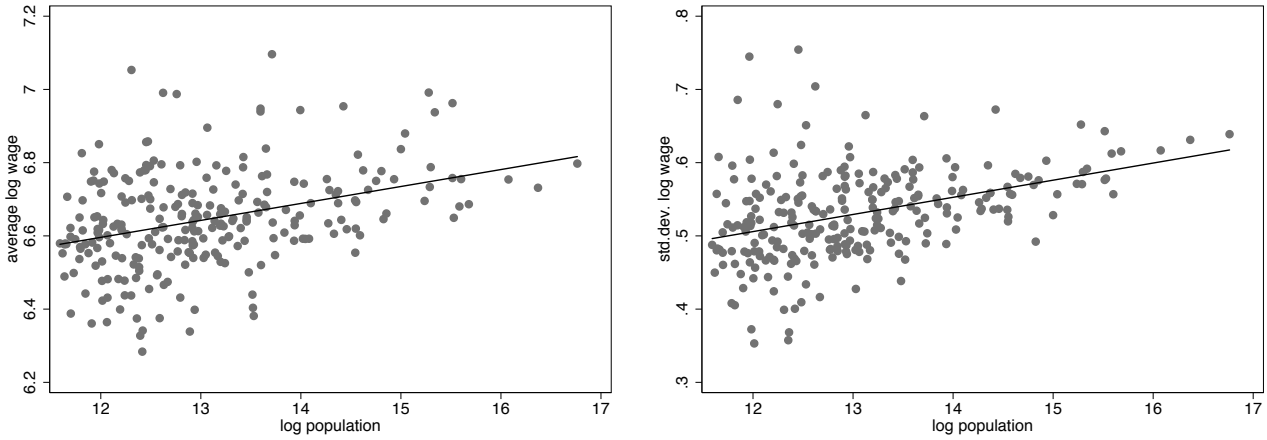

Figure 4: Wage distribution by population size. A. Mean (slope average=0.046 (s.e.=0.007); B. Standard Deviation (slope st.dev.=0.023 (s.e.=0.003)).

wage distribution for each city with maximum likelihood assuming log normality, i.e. a tobit regression on a constant. Figure 4 plots these estimates against 2009 population size. We see that both average wages and the variance of wages increases with population size. A simple linear regression estimates a slope coefficient of 0.046 for mean log wages (robust $t = 7.07$, $p > 0.000$) and of 0.023 (robust $t = 8.03$, $p > 0.000$) for the standard deviation of log wages. On average, a one percent increase in the city population leads to 0.046% increase in the wage. Table 2 shows the top 10 and bottom 10 cities with respect to average wages.

[Insert Table 2 here]

5.4 Housing Prices

We model housing as a homogenous good h with a location specific per unit price p_j. In practice, however, housing differs in many observable dimensions. Observed housing prices therefore reflect both the location and the physical characteristics of the unit. Sieg et al. (2002) show the conditions under which housing can be treated as if it were homogenous and how to construct a price index for it. Take our Cobb-Douglas utility function

$$u(c, h(z)) = c^{1-\alpha} h^{\alpha}(z)$$

and assume that housing $h(z)$ is a function, for simplicity of exposition only, of two characteristics $z = (z_1, z_2)$ with a nested Cobb-Douglas structure

$$h(z) = z_1^{\delta} z_2^{1-\delta}.$$

The indirect utility given the market prices q_1 and q_2 for, respectively, characteristic z_1 and z_2 is then

$$U_i = \alpha^{\alpha} (1-\alpha)^{1-\alpha} \left[L q_1^{\delta} q_2^{1-\delta} \right]^{-\alpha} w$$

where $L = 1/[\delta^{\delta} (1-\delta)^{1-\delta}]$. Defining the price index $p = L q_1^{\delta} q_2^{1-\delta}$ the indirect utility is

$$U_i = \alpha^{\alpha} (1-\alpha)^{1-\alpha} \frac{w}{p^{\alpha}}$$

and thus identical to the one derived assuming homogenous housing h with market price p. The sub-expenditure function $e(q_1, q_2, h)$ is defined as the minimum expenditure necessary to obtain h units of housing and given by

$$e(q_1, q_2, h) = L q_1^{\delta} q_2^{1-\delta} h = ph = p z_1^{\delta} z_2^{1-\delta}.$$

Taking logarithms and assuming that we observe z_1 but not z_2 yields a linear hedonic regression model

$$log(e_{jn}) = log(p_j) + \delta log(z_{1jn}) + u_{jn}$$

where e_n is the observed rental price of housing unit n and $log(p_j)$. We can therefore estimate the city specific price level as location-specific fixed effect in a simple hedonic regression of log rental prices on the physical characteristics.

[Insert Table 3 here]

Table 3 shows the results of the hedonic regressions both for rental units and owner-occupied units using Census data. We use all available housing characteristics in the data and add all categories as dummy variables without functional form assumptions. All coefficients are highly significant with expected signs: housing prices increase with the number of rooms and decrease with the age of the

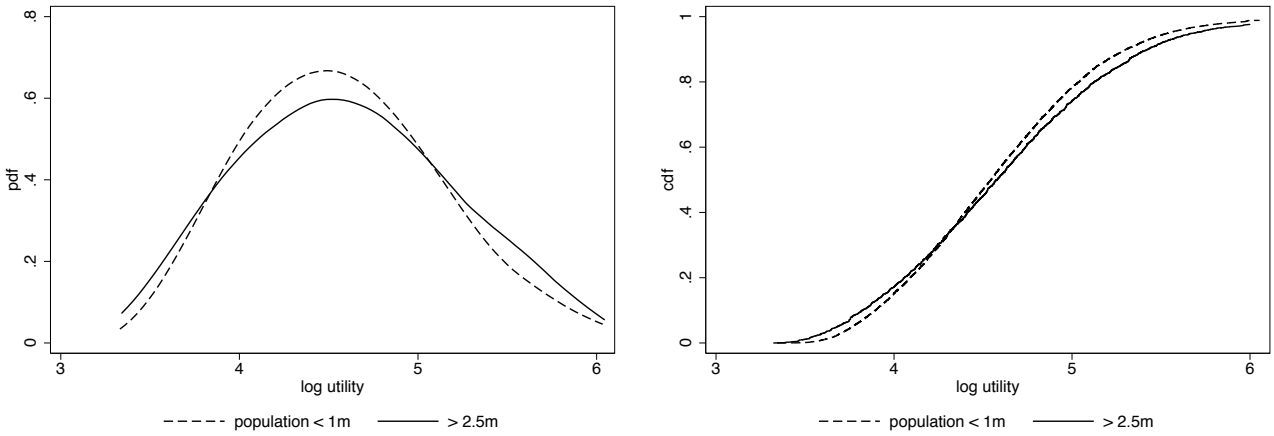

Figure 5: Skill distribution for small and large cities. A. Kernel density estimtates (Epanechnikov kernel, bandwidth = 0.2), *not* adjusted for city-specific top-coding; B. Empirical CDF adjusted for top-coding using the Kaplan-Meier method.

structure. We find a non-monotonic relationship in the numbers of units in the structure with highest prices for single-family detached homes and buildings with more than 50 units.

We adjust our estimated price levels from the 2000 Census for the 2000-2009 price changes using data from the Federal Housing Finance Agency (FHFA). Table 4 shows the resulting house price indices for the highest and lowest priced cities in our sample.

[Insert Table 4 here]

5.5 Skill distribution

Davis and Ortalo-Magné (2007) document that expenditure shares on housing are very constant across U.S. metropolitan areas with a median expenditure share of 0.24. We use this as our estimate of α. Together with our estimate for local housing prices p_j we can back out the indirect utility u_{ij} for the observed wages using equation (4).

Figure 5 shows the distribution of skills for full-time earners both in cities with a population of more than 2.5 million and cities with population between 100,000 and 1 million. In contrast to the wage distribution, the skill distribution in large cities is only marginally shifted to the right. However, both the upper and the lower tail of the distribution is thicker in the large cities thus confirming the theoretical prediction of fat tails.

The above partitioning of skills into a group of small cities and a group of large cities ignores substantial differences across different cities of similar size. We therefore estimate the mean and the standard deviation of the skill distribution for each city. As with wages, we take into account the city-specific right censoring from top-coded wages by estimating a censored (tobit) regression on a constant.

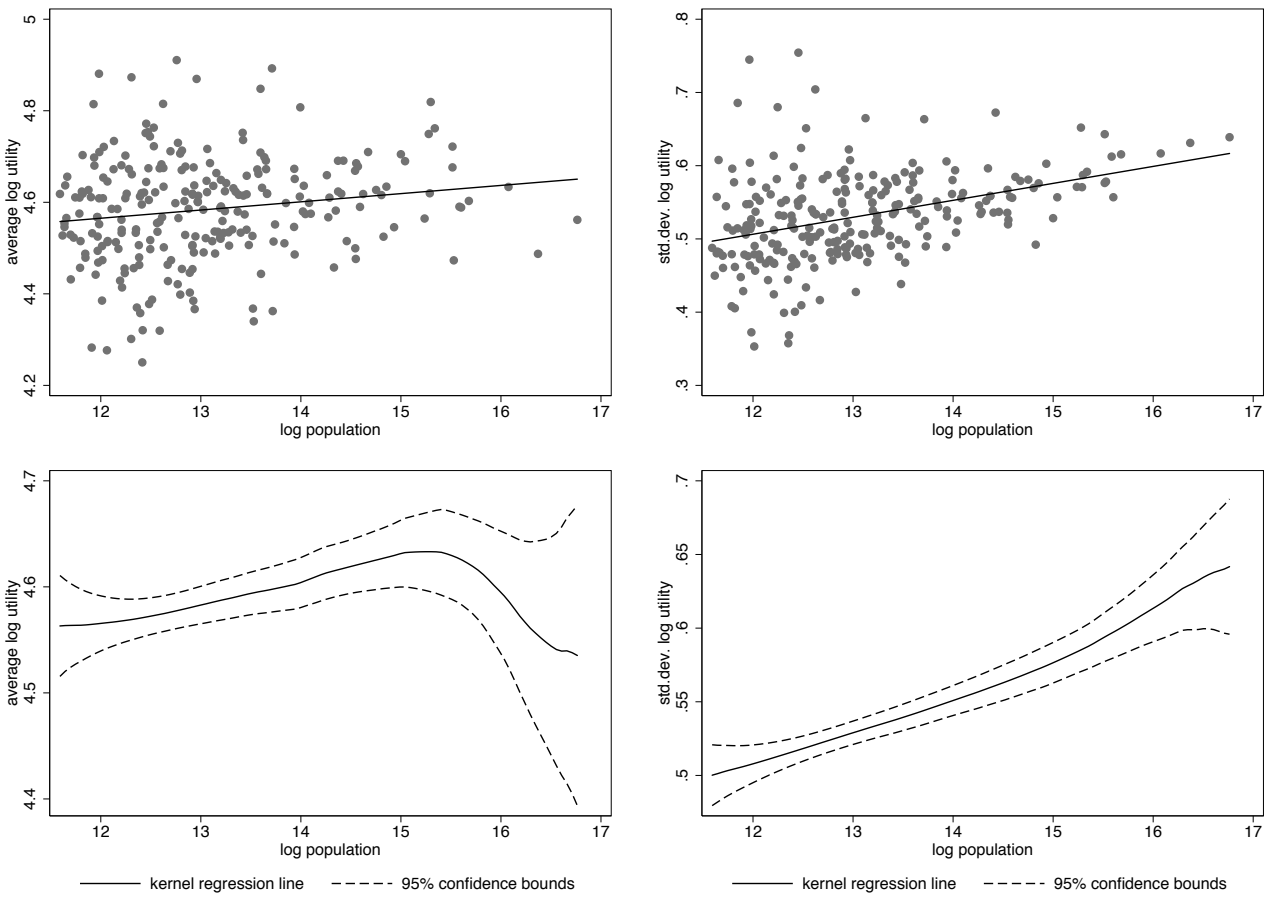

Figure 6: Skill distribution by population size. Left graphs: Mean; Right graphs: Standard Deviation. Top graphs: linear regression (slope average=0.018 (s.e.=0.006); slope st.dev.=0.023 (s.e.=0.003)). Bottom graphs: local linear regression (Epanechnikov kernel, bandwidth = 0.6 and 0.83, respectively)

The top two graphs in Figure 6 plot these estimates against 2009 population size. We see that while average skills vary little with population size, the standard deviation increases substantially. A simple linear regression estimates a slope coefficient of 0.018 for mean log utility (robust $t = 2.90$, $p > 0.004$) and of 0.023 (robust $t = 7.79$, $p > 0.000$) for the standard deviation of log utility. The lower two graphs in Figure 6 show non-parametric local linear regressions for the size relationship and 95% confidence intervals. Both the parametric and non-parametric estimates clearly confirm the fat tail hypothesis. Table 5 shows the top 10 and bottom 10 cities with respect to average wages.

[Insert Table 5 here]

[Insert Table 6 here]

6 Other Measures of Skills

As a robustness check and as external validation, we compare our implicit skill distribution with observed measures of skill. Figure 7 shows the distribution of educational attainment for the same CPS population as our wage data. The same pattern as with our implicit measure arises: both the highest and the lowest skilled workers are disproportionately more frequent in larger cities than in smaller ones.

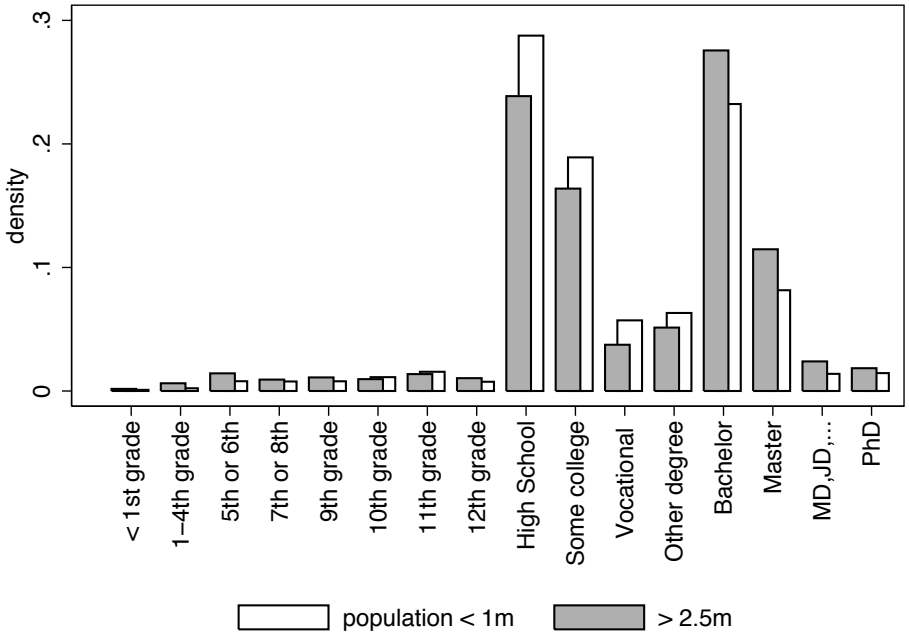

Figure 7: Observed educational attainment for small and large cities. Highest completed grade of full-time wage earner in 2009 CPS.

This can be observed even more transparently when we group the education levels into three groups. This is reported in Figure 8. What is most striking about this observation is that the fat tails in the distribution of educational attainment is obtained *independently* of how we obtained our measure of skills before. No theory is needed and the measure of skills is determined exogenously.

The fat tails in the distribution of educational attainment in larger cities can also be established at the individual city level. Below in Figure 9, we report the scatter plot of the variance of educational attainment when educational attainment categories are given a score corresponding to the years of schooling.

Like in the case where the skill measure is derived from the wage distribution, when we use an observable, reported measure of skill, we find little correlation between city size and average skill, but a significant and positive relation between city size and the standard deviation of the skill measure.

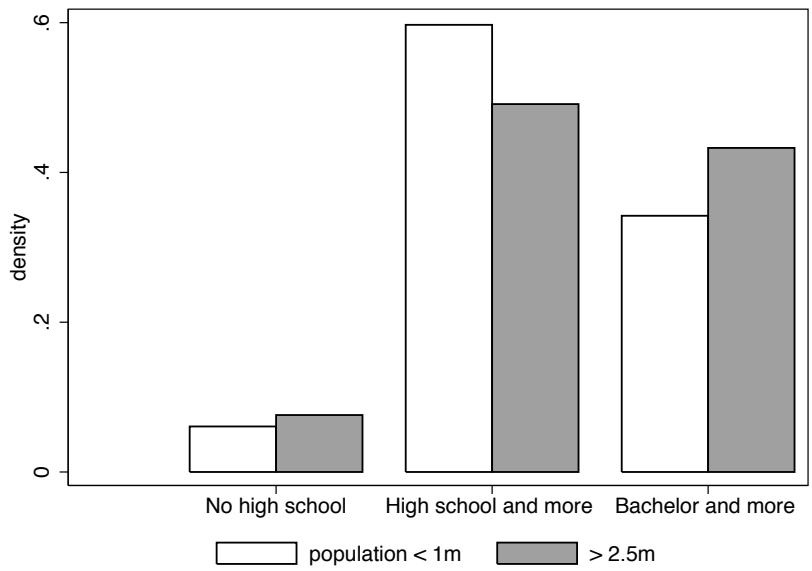

Figure 8: Observed educational attainment for small and large cities, grouped in 3 categories. Highest completed grade of full-time wage earner in 2009 CPS.

7 Discussion and Extensions

7.1 Firm size

In our model, firm size is endogenous. We can therefore identify primitive parameters from the empirical firm size distribution. We use Census data[8] on the number of employees and establishments for counties or CBSAs. This allows us to calculate the average number of employees per establishment by city,

Figure 10 reports the average firm size by city size. The linear regression coefficient is positive and positive. The kernel estimate is inverted U-shaped, though the downward sloping portion is not significant. In terms of the magnitude, the average firm size increases between 15 and 17 employees, from the kernel estimate.[9]

We can exploit the fact that theory pins down the relation between TFP and house prices. From Lemma 1 in the Appendix, we know that $m_{i1} > m_{i2} \iff \frac{A_1}{A_2} > \left(\frac{p_1}{p_2}\right)^{\alpha}$. This therefore implies that the ratio of TFP between two cities can be bounded by:

$$\frac{A_1}{A_2} > \left(\frac{p_1}{p_2}\right)^{\alpha} = \frac{w_{i1}}{w_{i2}},$$

[8] County Business Patterns, U.S. Census: http://www.census.gov/econ/cbp/index.html.

[9] For the service sector, Holmes and Stevens (2003) find a positive relation between city size and establishment size, and a negative relation in manufacturing. Given the modest size of the manufacturing sector (9% of all non-farm employment – www.bls.gov) relative to services (69%), this is consistent with our finding that across all sectors this relation is increasing.

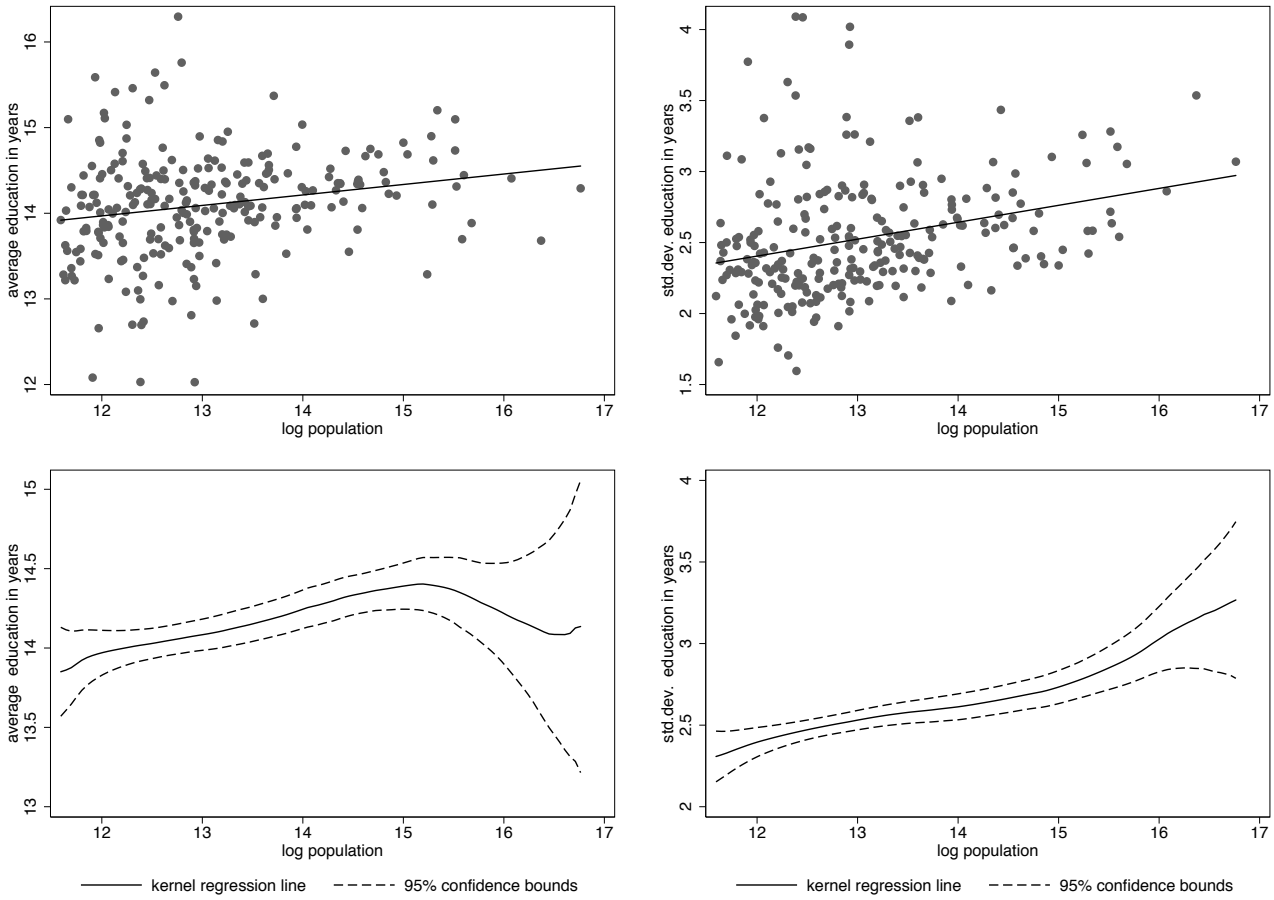

Figure 9: Distribution of educational attainment (translated into years) by population size. Left graphs: Mean; Right graphs: Standard Deviation. Top graphs: linear regression (slope average=0.12 (s.e.=0.03); slope st.dev.=0.12 (s.e.=0.02)). Bottom graphs: local linear regression (Epanechnikov kernel, bandwidth = 0.53 and 0.68, respectively)

where use the equilibrium condition of mobility across cities that the wage ratio must be proportional to the price ratio. TFP in the largest cities in our sample is at least 25% higher than that in the smallest cities (with a population around 160,000). The fact that the TFP is larger than labor productivity is due free entry of firms and the fact that the cost of entry depends on the house price index and is therefore different across cities.

7.2 The Role of Migration

Casual observation suggests that large cities tend to have a disproportionate representation of low skilled immigrant workers. Often kitchen staff in restaurants or construction workers are immigrants with low skills and incomes. And indeed, while foreign borns are overall a relatively small fraction of the working population (less than 10%), the data confirms that they are much more likely to locate

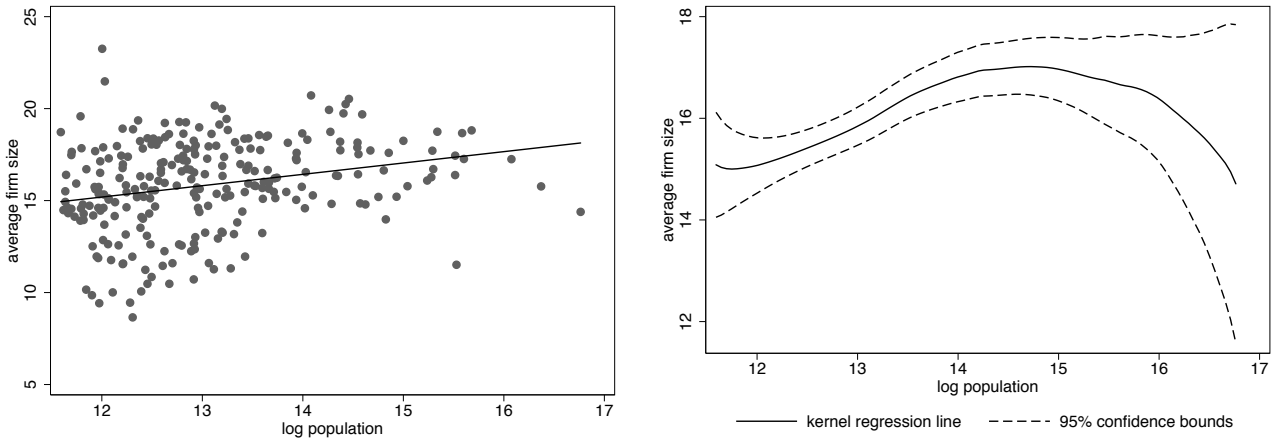

Figure 10: Average firm size by city population: A. linear regression (slope=0.62 (s.e.=0.14)); B. local linear regression (Epanechnikov kernel, bandwidth=0.58).

in large cities (12% of the work force) than in small cities (5%). Maybe the effect of disproportionate representation of the low skilled in large cities is driven by immigration.

In the context of our model it does not matter whether it is the low skilled Americans or low skilled immigrants who disproportionately locate in large cities. In equilibrium they should be indifferent. Of course, there is likely to be within-skill heterogeneity (in preferences for example), and some low skilled workers will strictly prefer to locate in either large or small cities. While we do not model this, in equilibrium there should still be arbitrage by the marginal worker within a skill type. Thus it may well be the case that migrants have certain benefits from locating in large cities. For example networks (see Munshi (2003)) play an important role for the location decision of migrants, and if only migrants have that benefit, at a competitively set wage, migrants will strictly prefer to locate in the city that offers the same utility plus the network benefit. Alternatively, migrants may locate in large cities due to limited information about smaller cities.

In any event, because even with those additional benefits for migrants, or any within skill heterogeneity, the model still predicts that in equilibrium, low skilled workers disproportionally move into large cities. It is sufficient that the marginal type within a skill class arbitrages the difference.

To evaluate the role of migrants in the location decision, we split the sample up into natives and foreign born workers. Figure 11 reports the plot of both distributions. Not surprisingly, the implied skill distribution for the foreign born is more skewed to the left than that of the natives. We find that even the distribution of foreign born workers has fat tails, *both* for the low *and* the high skilled. The latter is maybe most surprising: not only do the low skilled foreign born disproportionately migrate to large cities, so do the high skilled migrants. Most importantly, even after subtracting all the migrants, the distribution of natives has fatter tails in large cities. The fat tails are therefore not exclusively driven by non-natives.

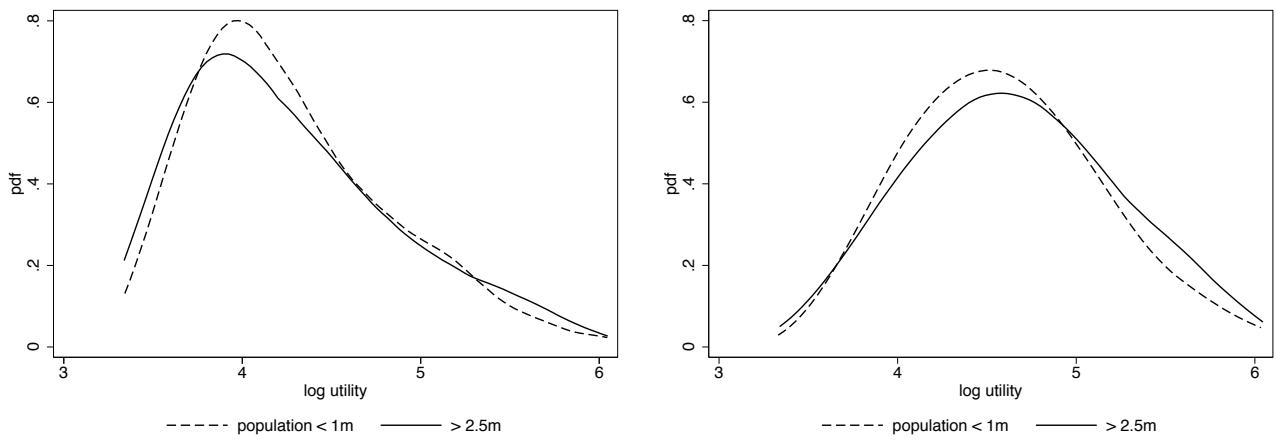

Figure 11: Skill distribution: A. Foreign born workers; B. Natives.

7.3 A Micro-foundation for VES: Spillovers from Skill Diversity

So far we have been agnostic about what determines the VES technology. It is well documented that agglomeration externalities are important (see for example Davis, Fisher and Whited (2009) among many others). Here we propose a simple micro-foundation for the technology with varying elasticities of substitution that generates the fat tails, and that is derived from spillovers across skill types.

The production technology in a city j is given by:

$$Y_j = A_j \sum_i a(\cdot) m_{ij}^{\gamma} y_i^{\beta}.$$

This technology is completely standard CES except for the fact that there is a knowledge spillover $a(\cdot) = m_{ij}^{\chi(\cdot)}$ that affects the marginal productivity of the worker. Knowledge spillovers are generated by the input of diversely skilled workers. Having a different viewpoint helps solve a problem (e.g., the input from the baggage loader at Southwest airlines on the performance of the logistics manager to streamline luggage flows). There is no spillover from meeting a same skilled type as that knowledge is already embodied in your own skill. We assume that spillovers arise whenever individuals meet, which occurs through uniform random matching. So if a worker meets one other worker per period, the probability that she is of another skill type is given by:

$$1 - \frac{M_i}{\sum_i M_i},$$

and the effect of the spillover on the marginal productivity then is

$$\chi \left(1 - \frac{M_i}{\sum_i M_i} \right)$$

and increasing. The nature of the spillover is illustrated in Figure 12.

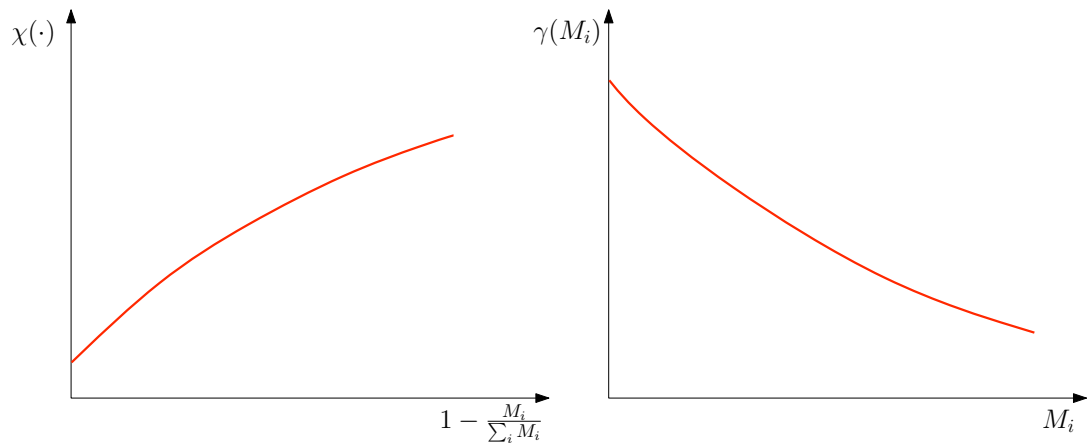

Figure 12: A. Spillover technology $\chi(\cdot)$, increasing in measure of other skills; B. The marginal product $\gamma(m_{ij})$ (and the Elasticity of Substitution ρ) are decreasing in abundance of skill.

Output for each skill now consists of

$$A_j m_{ij}^{\chi\left(1-\frac{M_i}{\sum_i M_i}\right)} m_{ij}^{\gamma} y_i^{\beta}$$

and introducing the notation $\gamma(M_i) = \chi\left(1 - \frac{M_i}{\sum_i M_i}\right) + \gamma$ where $\gamma(M_i)$ is a decreasing function, we can write the technology as

$$Y_j = A_j \sum_i m_{ij}^{\gamma(M_i)} y_i^{\beta}.$$

Irrespective of the functional form of $\gamma(\cdot)$, the important implication of this formulation of the technology is that it is a variation on the standard CES technology, except for the fact that the elasticity varies by skill. This Varying Elasticity of Substitution (VES) technology of course is no longer homothetic. There is still a direct relation between $\gamma(\cdot)$ and the elasticity of substitution ρ_{ik} between skill i and k is given by (see Appendix for the derivation):

$$\rho ij = \frac{\gamma_i n_i^{\gamma_i} y_i + \gamma_j n_j^{\gamma_j} y_j}{\gamma_i n_i^{\gamma_i} y_i \left(1 - \gamma_j\right) + \gamma_j n_j^{\gamma_j} y_j \left(1 - \gamma_i\right)},$$

where $\gamma_i = \gamma(M_i)$ and $\gamma_k = \gamma(M_k)$. Observe that if $\gamma_i = \gamma_k = \gamma$, the technology is CES and this expression simplifies to the usual constant elasticity $\rho = \frac{1}{1-\gamma}$. The technology with varying elasticity is compared to the constant elasticity technology in Figure 13.

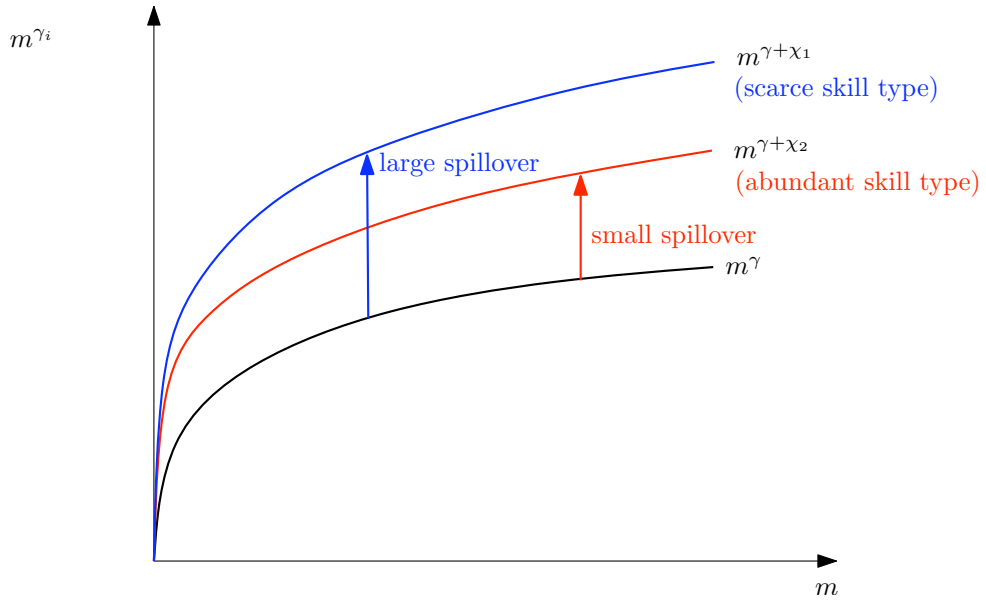

Figure 13: Scarce skill types are more likely to interact with agents with different skills and therefore, given m, they benefit from a larger expected spillover.

7.4 Unemployment

One alternative explanation for the fat tails may emanate from market frictions (see for example Eeckhout and Kircher (2010), Eeckhout, Lentz and Roys (2010), Gautier, Svarer and Teulings (2010) and Helpman, Itskhoki and Redding (2010)). Consider a CES technology but with search frictions. Then more abundant skill types will face a relatively high unemployment to vacancy ratio, whereas scarce skill types face a low ratio. This drives a wedge between the marginal productivity and wages. In a labor market without an urban dimension, Eeckhout, Lentz and Roys (2010) show in a directed search model that this leads to fatter tails in the more productive firms. It remains to be verified though that empirically the unemployment rate both for high and low skilled workers is substantially higher.

8 Conclusion

We have proposed a tractable theory of spatially dispersed production with perfectly mobile heterogeneous inputs, skilled labor. Differences in TFP lead to differences in demand for skills across cities. In general equilibrium, wages and house prices clear the labor and housing markets. Perfect mobility of citizens leads to utility equalization by skill.

We show that cities with a higher TFP are larger and that a CES production technology entails identical skill distributions across cities with different productivity. When the elasticity of substitution varies across skills such that it is higher for scarce skills, the skill distribution in larger cities exhibits fatter tails.

We find empirical support for our theory using US data. Adjusting wages for the compensating differentials of house prices by means of a hedonic price index, we find skill distributions that have fatter tails in larger cities. Our measure of skill derives directly from wages, and includes therefore also unobservable determinants of skills. For external validation, we also use a measure of observable skills only – years of schooling – and find the same results. Of course, in order to capture the non-monotonic relation in the demand for skills, the partition of skill classes must be sufficiently fine. The robustness of the result to the use of skill measures based on both observables and unobservables and measures based on observables only is indicative of the robustness of the result. The use of a wage based skill measure is not only attractive because it incorporates unobservable characteristics of skill, by construction it is also measured as a continuous variable. While partitioning worker types in two classes of high and low skilled is attractive in many ways, it precludes identification of non-linear relations, let alone non-monotonic relations.

Appendix A: Theory

Proof of Theorem 1

Proof. Given constant γ, we can rewrite the first equilibrium condition as:

$$\frac{\sum_{i=1}^{I}\left(\frac{M_i}{N_2}\right)^{\gamma}\gamma y_i^{\beta}\left\{\frac{\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma-1}}}{1+\frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma-1}}}\right\}^{\gamma}}{\sum_{i=1}^{I}\left(\frac{M_i}{N_2}\right)^{\gamma}\gamma y_i^{\beta}\left\{\frac{1}{1+\frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma-1}}}\right\}^{\gamma}}=\frac{H_1}{H_2}\frac{A_2}{A_1}\frac{N_2}{N_1}\frac{p_1}{p_2}$$

and therefore after canceling common terms as:

$$\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma}{\gamma-1}}=\frac{H_1}{H_2}\frac{A_2}{A_1}\frac{N_2}{N_1}\frac{p_1}{p_2},$$

We solve for the price ratio:

$$\frac{p_1}{p_2}=\left(\frac{H_1}{H_2}\frac{N_2}{N_1}\right)^{\frac{\gamma-1}{1-\gamma(1-\alpha)}}\left(\frac{A_2}{A_1}\right)^{-\frac{1}{1-\gamma(1-\alpha)}}.$$

Observe that:

$$\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}=\left[\left(\frac{H_1}{H_2}\frac{N_2}{N_1}\right)^{\alpha}\left(\frac{A_2}{A_1}\right)^{\alpha-1}\right]^{\frac{\gamma-1}{1-\gamma(1-\alpha)}}$$

Substituting into the expressions for m_{ij} we obtain:

$$\begin{aligned}
m_{i1} &= \frac{\left[\left(\frac{H_1}{H_2}\frac{N_2}{N_1}\right)^{\alpha}\left(\frac{A_2}{A_1}\right)^{\alpha-1}\right]^{\frac{1}{1-\gamma(1-\alpha)}}}{N_2+N_1\left[\left(\frac{H_1}{H_2}\frac{N_2}{N_1}\right)^{\alpha}\left(\frac{A_2}{A_1}\right)^{\alpha-1}\right]^{\frac{1}{1-\gamma(1-\alpha)}}}M_i\\
m_{i2} &= \frac{1}{N_2+N_1\left[\left(\frac{H_1}{H_2}\frac{N_2}{N_1}\right)^{\alpha}\left(\frac{A_2}{A_1}\right)^{\alpha-1}\right]^{\frac{1}{1-\gamma(1-\alpha)}}}M_i
\end{aligned}$$

The density at any skill level i is simply the ratio of the measure of that skill over the total measure,

and after simplifying, we get:

$$\frac{m_{i1}}{\sum_{i=1}^{I} m_{i1}} = \frac{\left(\frac{\left[\left(\frac{H_1}{H_2} \frac{N_2}{N_1} \right)^{\alpha} \left(\frac{A_2}{A_1} \right)^{\alpha-1} \right]^{\frac{1}{1-\gamma(1-\alpha)}}}{N_2 + N_1 \left[\left(\frac{H_1}{H_2} \frac{N_2}{N_1} \right)^{\alpha} \left(\frac{A_2}{A_1} \right)^{\alpha-1} \right]^{\frac{1}{1-\gamma(1-\alpha)}}} \right) M_i}{\left(\frac{\left[\left(\frac{H_1}{H_2} \frac{N_2}{N_1} \right)^{\alpha} \left(\frac{A_2}{A_1} \right)^{\alpha-1} \right]^{\frac{1}{1-\gamma(1-\alpha)}}}{N_2 + N_1 \left[\left(\frac{H_1}{H_2} \frac{N_2}{N_1} \right)^{\alpha} \left(\frac{A_2}{A_1} \right)^{\alpha-1} \right]^{\frac{1}{1-\gamma(1-\alpha)}}} \right) \sum_{i=1}^{I} M_i} = \frac{M_i}{\sum_{i=1}^{I} M_i}$$

Likewise for the density in city 2:

$$\frac{m_{i2}}{\sum_{i=1}^{I} m_{i2}} = \frac{M_i}{\sum_{i=1}^{I} M_i}$$

Therefore, both distributions are identical and equal to the economy-wide distribution. ■

Proof of Proposition 1

First we prove the following Lemma concerning the housing prices.

Lemma 1 *When* $\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} < 1$, $m_{i1} > m_{i2}$, $\forall i \in I$.

Proof. Recall that we have $A_1 > A_2$. Defining $Z = \frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha}$. From $Z < 1$, we know that $Z^{\frac{1}{\gamma_i - 1}} > 1$, since $\gamma_i \in (0, 1)$. Then, from the first order conditions, we obtain:

$$m_{i1} = \frac{Z^{\frac{1}{\gamma_i - 1}}}{N_2 + N_1 Z^{\frac{1}{\gamma_i - 1}}} M_i > \frac{1}{N_2 + N_1 Z^{\frac{1}{\gamma_i - 1}}} M_i = m_{i2}$$

■

Now, we prove the Proposition:

Proof. Consider the system of equilibrium equations (1)–(3), with $H_1 = H_2$. Equating (2) and (3), we obtain:

$$\sum_{i=1}^{I} \left(\frac{M_i}{N_2} \right)^{\gamma_i} \frac{A_2}{p_2} \frac{(1-\gamma_i)\, y_i^{\beta}}{\left(1 + \frac{N_1}{N_2} \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_i - 1}} \right)^{\gamma_i}} \left\{ \frac{A_1}{A_2} \left(\frac{p_2}{p_1} \right) \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{\gamma_i}{\gamma_i - 1}} - 1 \right\} = 0 \tag{5}$$

and after rearranging, we have:

$$\frac{A_1}{A_2} \left(\frac{p_2}{p_1} \right) \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{\gamma_i}{\gamma_i - 1}} = \left(\frac{A_1}{A_2} \right)^{\frac{1}{1-\gamma_i}} \left(\frac{p_2}{p_1} \right)^{1 + \frac{\alpha \gamma_i}{1-\gamma_i}}.$$

Since $\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma_i}} > 0$, it immediately follows that $p_2 < p_1$.

The term inside curly brackets can be written as:

$$\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_i}{\gamma_i-1}} - 1 = \left(\frac{p_2}{p_1}\right)^{1-\alpha}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}} - 1,$$

and given $\frac{p_2}{p_1} < 1$, the equality in equation (5) requires that $\left(\frac{p_2}{p_1}\right)^{1-\alpha}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}} \geq 1$ for some values of γ_i. This is only satisfied if $\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}} > 1$ for some values of γ_i. But this is only possible if $\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha} < 1$. Therefore, $Z < 1$.

From Lemma 1, this imply that $m_{i1} > m_{i2}$. Therefore, each individual firm in city 1 is bigger than each individual firm in city 2.

The economy can be fully characterized by the system of equations:

$$\sum_{i=1}^{N} \alpha \frac{w_{i1}}{p_1} m_{i1} = \frac{H_1}{N_1} \tag{6}$$

$$\sum_{i=1}^{N} \alpha \frac{w_{i2}}{p_2} m_{i2} = \frac{H_2}{N_2} \tag{7}$$

$$m_{i1} = \left(\frac{w_{i1}}{\gamma_i A_1 y_i}\right)^{\frac{1}{\gamma_i-1}}$$

$$m_{i2} = \left(\frac{w_{i2}}{\gamma_i A_2 y_i}\right)^{\frac{1}{\gamma_i-1}}$$

$$N_1 m_{i1} + N_2 m_{i2} = M_i$$

$$\frac{w_{i1}}{(p_1)^{\alpha}} = \frac{w_{i2}}{(p_2)^{\alpha}}$$

$$A_1 \sum_{i=1}^{N} (1-\gamma_i)(m_{i1})^{\gamma_i} y_i = kp_1$$

$$A_2 \sum_{i=1}^{N} (1-\gamma_i)(m_{i2})^{\gamma_i} y_i = kp_2$$

From $\frac{w_{i1}}{(p_1)^{\alpha}} = \frac{w_{i2}}{(p_2)^{\alpha}}$, $\forall i \in \{1, ..., N\}$, and $p_1 > p_2$ we have that $w_{i1} > w_{i2}$. Now, consider that:

$$\frac{w_{ij}}{p_j} = \frac{1}{(p_1)^{1-\alpha}} \frac{w_{i1}}{(p_1)^{\alpha}} = \frac{1}{(p_1)^{1-\alpha}} \frac{w_{i2}}{(p_2)^{\alpha}}$$

$$\frac{w_{i2}}{p_2} = \frac{1}{(p_2)^{1-\alpha}} \frac{w_{i2}}{(p_2)^{\alpha}}$$

Then:

$$\frac{w_{i1}}{p_1} = \left(\frac{p_2}{p_1}\right)^{1-\alpha} \frac{w_{i2}}{p_2} < \frac{w_{i2}}{p_2},$$

since $\frac{p_2}{p_1} < 1$.

Then, from equations (6) and (7), we obtain, for $H_1 = H_2$:

$$\sum_{i=1}^{N} \alpha \left[N_1 \frac{w_{i1}}{p_1} m_{i1} - N_2 \frac{w_{i2}}{p_2} m_{i2} \right] = 0$$

Rearranging, we have:

$$\sum_{i=1}^{N} \alpha \frac{w_{i2}}{p_2} m_{i2} \left[N_1 \left(\frac{p_2}{p_1} \right)^{1-\alpha} \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_i - 1}} - N_2 \right] = 0$$

Notice that:

$$\frac{dZ^{\frac{1}{\gamma_i - 1}}}{d\gamma_i} = -\frac{Z^{\frac{1}{\gamma_i - 1}}}{(\gamma_i - 1)^2} \ln Z.$$

Since $Z < 1$, we have that $\frac{dZ^{\frac{1}{\gamma_i - 1}}}{d\gamma_i} > 0$. This implies that the first term inside squared brackets is increasing in γ_i. Since it is definitely positive for $\gamma_i \approx 1$, it must be negative around $\gamma_i = 0$. Therefore, we have that:

$$N_1 \left(\frac{p_2}{p_1} \right)^{1-\alpha} \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{-1} - N_2 \quad < \quad 0$$

Rearranging this implies:

$$\frac{N_1}{N_2} < \frac{A_2}{A_1} \frac{p_1}{p_2}$$

Now we can compare the size of cities 1 and 2 :

$$S_2 - S_1 \quad = \quad \sum_{i-1}^{N} N_2 m_{i2} \left[1 - \frac{N_1}{N_2} \frac{m_{i1}}{m_{i2}} \right]$$

Since $\frac{m_{i1}}{m_{i2}} = \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_i - 1}}$ and $\frac{N_1}{N_2} < \frac{A_2}{A_1} \frac{p_1}{p_2}$, we have:

$$S_2 - S_1 < \sum_{i-1}^{N} N_2 m_{i2} \left[1 - \left(\frac{A_2}{A_1} \frac{p_1}{p_2} \right) \times \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_i - 1}} \right]$$

We know that

$$\left(\frac{A_2}{A_1} \frac{p_1}{p_2} \right) \times \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{1}{\gamma_i - 1}} = \left(\frac{p_1}{p_2} \right)^{1-\alpha} \times \left[\frac{A_2}{A_1} \left(\frac{p_1}{p_2} \right)^{\alpha} \right]^{\frac{\gamma_i}{\gamma_i - 1}} > 1.$$

Therefore $S_1 > S_2$. ■

8.1 Proof of Theorem 2

Proof. Denote the mode of the skill distribution by $\bar{i}$. Then the distribution is uni-modal if for all $i' > i$, $M_{i'} > M_i$ when $i, i' \leq \bar{i}$ and $M_{i'} < M_i$ when $i, i' \geq \bar{i}$.

From the assumptions in the theorem and by the proof of Proposition 1, we know that $Z < 1$ and therefore $Z^{\frac{1}{\gamma_i - 1}} > 1$. The density of any skill i in each city is given by

$$\frac{m_{i1}}{\sum_{k=1}^{I} m_{i1}} = \frac{\frac{Z^{\frac{1}{\gamma_i - 1}}}{N_2 + N_1 Z^{\frac{1}{\gamma_i - 1}}} M_i}{\sum_{k=1}^{I} \frac{Z^{\frac{1}{\gamma_k - 1}}}{N_2 + N_1 Z^{\frac{1}{\gamma_k - 1}}} M_k}$$

and

$$\frac{m_{i2}}{\sum_{k=1}^{I} m_{k2}} = \frac{\frac{1}{N_2 + N_1 Z^{\frac{1}{\gamma_i - 1}}} M_i}{\sum_{k=1}^{I} \frac{1}{N_2 + N_1 Z^{\frac{1}{\gamma_k - 1}}} M_k}.$$

Therefore the ratio of densities is

$$\frac{\frac{m_{i1}}{\sum_{k=1}^{I} m_{k1}}}{\frac{m_{i2}}{\sum_{k=1}^{I} m_{k2}}} = \frac{Z^{\frac{1}{\gamma_i - 1}} \sum_{k=1}^{I} \frac{1}{N_2 + N_1 Z^{\frac{1}{\gamma_k - 1}}} M_k}{\sum_{k=1}^{I} \frac{Z^{\frac{1}{\gamma_k - 1}}}{N_2 + N_1 Z^{\frac{1}{\gamma_k - 1}}} M_k}.$$

First, we write the ratio of densities in both cities for the highest skills $i = I$ (given a symmetric distribution, exactly the same holds for $i = 1$):

$$\frac{\frac{m_{I1}}{\sum_{k=1}^{I} m_{k1}}}{\frac{m_{I2}}{\sum_{k=1}^{I} m_{k2}}} = \frac{Z^{\frac{1}{\gamma_I - 1}} \sum_{k=1}^{I} \frac{1}{N_2 + N_1 Z^{\frac{1}{\gamma_k - 1}}} M_k}{\sum_{k=1}^{I} \frac{Z^{\frac{1}{\gamma_k - 1}}}{N_2 + N_1 Z^{\frac{1}{\gamma_k - 1}}} M_k} > 1.$$

The inequality follows from the fact that $\gamma_I < \gamma_k, \forall k \neq I$ and $Z < 1$ so that $Z^{\frac{1}{\gamma_I - 1}} > Z^{\frac{1}{\gamma_k - 1}}, \forall k \neq I$. It now also becomes clear that for a small enough grid of skills (i.e., I large), this inequality will also hold for skills in the neighborhood of $i = I$: $i = I - 1, I - 2, ...$

Now write the ratio of densities for $I = \bar{i}$:

$$\frac{\frac{m_{\bar{i}1}}{\sum_{k=1}^{I} m_{k1}}}{\frac{m_{\bar{i}2}}{\sum_{k=1}^{I} m_{k2}}} = \frac{Z^{\frac{1}{\gamma_{\bar{i}} - 1}} \sum_{k=1}^{I} \frac{1}{N_2 + N_1 Z^{\frac{1}{\gamma_k - 1}}} M_k}{\sum_{k=1}^{I} \frac{Z^{\frac{1}{\gamma_k - 1}}}{N_2 + N_1 Z^{\frac{1}{\gamma_k - 1}}} M_k} < 1.$$

Now the inequality follows from the fact that $\gamma_{\bar{i}} < \gamma_k, \forall k \neq \bar{i}$ and $Z < 1$ so that $Z^{\frac{1}{\gamma_{\bar{i}} - 1}} < Z^{\frac{1}{\gamma_k - 1}}, \forall k \neq I$. Again, the inequality will continue to hold even in a neighborhood of $\bar{i}$ provided the grid of skills is fine

enough.

As a result, the distribution in city 1, the high TFP city, has fatter tails and less density around the mode. From Proposition 1 we also know that city 1 - the high TFP city - is larger. Therefore the larger city has fatter tails. ■

Elasticity of Substitution

From the definition of the elasticity of substitution:

$$\sigma = \frac{\frac{d(x_2/x_1)}{x_2/x_1}}{\frac{d\left(\frac{df}{dx_1}/\frac{df}{dx_2}\right)}{\frac{df}{dx_1}/\frac{df}{dx_2}}}$$

Following Silberberg, the Elasticity of Substitution is given by:

$$\sigma = -\frac{f_1 f_2 \left(f_1 x_1 + f_2 x_2\right)}{x_1 x_2 \left(f_2^2 f_{11} - 2 f_1 f_2 f_{12} + f_1^2 f_{22}\right)}$$

In our case, we have:

$$\begin{aligned}
f_i &= A\gamma_i n_i^{\gamma_i - 1} y_i \\
f_j &= A\gamma_j n_j^{\gamma_j - 1} y_j \\
f_{ii} &= A\gamma_i \left(\gamma_i - 1\right) n_i^{\gamma_i - 2} y_i \\
f_{jj} &= A\gamma_j \left(\gamma_j - 1\right) n_j^{\gamma_j - 2} y_j \\
f_{ij} &= 0
\end{aligned}$$

Then, we have:

$$\begin{aligned}
f_i n_i + f_j n_j &= A\gamma_i n_i^{\gamma_i} y_i + A\gamma_j n_j^{\gamma_j} y_j \\
f_i f_j &= A^2 \gamma_i \gamma_j n_i^{\gamma_i - 1} n_j^{\gamma_j - 1} y_i y_j
\end{aligned}$$

Therefore, the numerator is:

$$A^3 \gamma_i \gamma_j y_i y_j n_i^{\gamma_i - 1} n_j^{\gamma_j - 1} \times \left(\gamma_i n_i^{\gamma_i} y_i + \gamma_j n_j^{\gamma_j} y_j\right)$$

and

$$\begin{aligned}
f_i^2 f_{jj} &= A^3 \gamma_i^2 n_i^{2\gamma_i - 2} y_i^2 \gamma_j \left(\gamma_j - 1\right) n_j^{\gamma_j - 2} y_j \\
f_j^2 f_{ii} &= A^3 \gamma_j^2 n_j^{2\gamma_j - 2} y_j^2 \gamma_i \left(\gamma_i - 1\right) n_i^{\gamma_i - 2} y_i
\end{aligned}$$

Then, the denominator is:

$$n_i n_j \times \left(\begin{array}{c} A^3 \gamma_i^2 n_i^{2\gamma_i - 2} y_i^2 \gamma_j \left(\gamma_j - 1\right) n_j^{\gamma_j - 2} y_j + \\ + A^3 \gamma_j^2 n_j^{2\gamma_j - 2} y_j^2 \gamma_i \left(\gamma_i - 1\right) n_i^{\gamma_i - 2} y_i \end{array} \right)$$

Rearranging, we have:

$$A^3 \gamma_i \gamma_j y_i y_j n_i^{\gamma_i - 1} n_j^{\gamma_j - 1} \times \Big(\gamma_i n_i^{\gamma_i} y_i \left(\gamma_j - 1\right) + \gamma_j n_j^{\gamma_j} y_j \left(\gamma_i - 1\right) \Big)$$

Therefore, the elasticity of substitution becomes:

$$\sigma = - \frac{A^3 \gamma_i \gamma_j y_i y_j n_i^{\gamma_i - 1} n_j^{\gamma_j - 1} \times \Big(\gamma_i n_i^{\gamma_i} y_i + \gamma_j n_j^{\gamma_j} y_j \Big)}{A^3 \gamma_i \gamma_j y_i y_j n_i^{\gamma_i - 1} n_j^{\gamma_j - 1} \times \Big(\gamma_i n_i^{\gamma_i} y_i \left(\gamma_j - 1\right) + \gamma_j n_j^{\gamma_j} y_j \left(\gamma_i - 1\right) \Big)}$$

Simplifying, we get:

$$\sigma = \frac{\gamma_i n_i^{\gamma_i} y_i + \gamma_j n_j^{\gamma_j} y_j}{\gamma_i n_i^{\gamma_i} y_i \left(1 - \gamma_j\right) + \gamma_j n_j^{\gamma_j} y_j \left(1 - \gamma_i\right)}$$

Notice that if $\gamma_i = \gamma_j = \gamma$, this expression simplifies to $\frac{1}{1-\gamma}$.

Appendix B: Data

Wage Data

Wage data is taken from the Current Population Survey (CPS), a joint effort between the Bureau of Labor Statistics (BLS) and the Census Bureau.[10] The CPS is a monthly survey and used by the U.S. Government to calculate the official unemployment and labor force participation figures. We the 2009 merged outgoing rotation groups (MORG) as provided by the National Bureau of Economic Research (NBER) [11]. The MORG are extracts of the basic monthly data during the household's fourth and eighth month in the survey, when usual weekly hours/earnings are asked.

We use the variable 'earnwke' as created by the NBER.[12] This variable reports earnings per week in the current job. It includes overtime, tips and commissions. For hourly workers, Item 25a ("How many hours per week does...usually work at this job?") times Item 25c ("How much does ...earn per hour?") appears here. For weekly workers, Item 25d ("How much does...usually earn per week at this job before deductions?") appears here.

We restrict the sample to full time workers (between 36 and 60 usual hours per week) with hourly wages above the federal minimum wage of 7.25 USD. Our final wage sample includes 102,599 workers out of the 320,941 surveyed persons. CPS wage data is in 2009 top-coded at a weekly wage of 2884.61 USD which applies to 2616 or 2.5% of workers. All estimations use the weights in variable 'earnwt' provided by the NBER.

The NBER version of the CPS identifies the core-based statistical area (CBSA) of the observation. It use the the New England city and town areas (NECTA) definition and codes for metro areas in the 6 New England states and the Federal Information Processing Standards (FIPS) definition and codes for all other states.

House Price Data

We use the 5% Public Use Micro Sample (PUMS) of the 2000 U.S. Census. The U.S. is a decennial random sample of housing units across the U.S. The data is provided by the Minnesota Population Center in its Integrated Public Use Microdata Series (IPUMS).[13]

The variable 'rent' reports the monthly contract rent for rental units and the variable 'valueh' the value of housing units in contemporary dollars. We also use all the reported housing characteristics of the unit: 'rooms' is the number of rooms, 'bedrooms' is the number of bedrooms, 'unitsstr' is the units in structure (in 8 groups), 'builtyr' is the age of structure (in 9 age groups) and 'kitchen' is a dummy variable if the unit has kitchen or cooking facilities.

[10] See http://www.bls.gov/cps/

[11] Stata data file available at http://www.nber.org/morg/annual/morg09.dta

[12] See details of the variable creation at the NBER website http://www.nber.org/cps/

[13] See Ruggles et al. (2010) for the data source and http://usa.ipums.org/usa/ for a detailed description of data and variables.

We drop housing units in group quarters, farmhouses, drop mobile homes, trailers, boats, and tents and only use data from housing units in identified metropolitan or micropolitan core based statistical areas (CBSA). Our final sample contains 3,274,198 rental and 7,680,728 owner occupied units.

The 5% PUMS discloses the co-called Public Use Microdata Area (PUMA). PUMA's are areas with a maximum of 179,405 housing units and only partly overlap with political borders of towns and counties. We use the Geographic Correspondence Engine with Census 2000 Geography from the Missouri Census Data Center(MCDC) [14] to link PUMA areas to CBSAs. The MCDC data matches every urban PUMA code to one or more CBSA codes and reports the fraction of housing units that are matched. We assign a PUMA to a CBSA if this fraction is bigger than 33%. In cases where the PUMA does not fully belong to a CBSA, we assign the PUMA to the CBSA where most of its housing units belong to. Our final sample contains data from 533 metropolitan or micropolitan core based statistical areas (CBSA) out of a total of 940 existing CBSAs. Not that we do *not* use the metropolitan area code provided in the PUMS in variable 'metaread'. This variable reports a mixture of metropolitan area codes (MSA, PMSA, central city or county) which is difficult to match with the CBSA definition.

We adjust our estimated price levels from the 2000 Census for the 2000-2009 price changes using data from the Federal Housing Finance Agency (FHFA).[15]. The FHFA publishes quarterly time series of local house price indices for 384 CBSAs based on the Case and Shiller (1987) repeat sales method. 11 of the CBSAs are divided into 29 metropolitan divisions. We average these divisions over the respective CBSA using the 2000 housing stock (provided by the MCDC) as weights.

[14] Available at http://mcdc2.missouri.edu/websas/geocorr2k.html.
[15] See http://www.fhfa.gov

Appendix C: Tables

Table 1: Rank of cities by 2009 population.

	City	Population
1	New York-Northern New Jersey-Long Island, NY-NJ-PA	19,069,796
2	Los Angeles-Long Beach-Santa Ana, CA	12,874,797
3	Chicago-Naperville-Joliet, IL-IN-WI	9,580,567
4	Dallas-Fort Worth-Arlington, TX	6,447,615
5	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	5,968,252
6	Houston-Sugar Land-Baytown, TX	5,867,489
7	Miami-Fort Lauderdale-Pompano Beach, FL	5,547,051
8	Washington-Arlington-Alexandria, DC-VA-MD-WV	5,476,241
9	Atlanta-Sandy Springs-Marietta, GA	5,475,213
10	Boston-Cambridge-Quincy, MA-NH	4,588,680
244	Albany, GA	165,440
245	Waterloo-Cedar Falls, IA	164,913
246	Panama City-Lynn Haven-Panama City Beach, FL	164,767
247	Oshkosh-Neenah, WI	163,370
248	Parkersburg-Marietta-Vienna, WV-OH	160,905
249	Niles-Benton Harbor, MI	160,472
250	Janesville, WI	160,155
251	Abilene, TX	160,070
252	Eau Claire, WI	160,018
253	Jackson, MI	159,828

Notes: cities are defined as core based statistical areas (CBSA). The Office of Management and Budget (OMB) defines 940 metropolitan and micropolitan areas of which we use the largest 253.

Table 2: Rank of cities by average log wages.

	City	Population	Avg. Log Wage
1	Bridgeport-Stamford-Norwalk, CT	901,208	7.10
2	Barnstable Town, MA	221,151	7.05
3	San Francisco-Oakland-Fremont, CA	4,317,853	6.99
4	Boulder, CO	303,482	6.99
5	Ann Arbor, MI	347,563	6.99
6	Washington-Arlington-Alex., DC-VA-MD-WV	5,476,241	6.96
7	San Jose-Sunnyvale-Santa Clara, CA	1,839,700	6.95
8	Worcester, MA	803,701	6.95
9	Hartford-West Hartford-East Hartford, CT	1,195,998	6.94
10	Oxnard-Thousand Oaks-Ventura, CA	802,983	6.94
244	Albany, GA	165,440	6.42
245	Columbia, SC	744,730	6.40
246	Johnson City, TN	197,381	6.40
247	Corpus Christi, TX	416,095	6.40
248	Bowling Green, KY	120,595	6.39
249	El Paso, TX	751,296	6.38
250	Utica-Rome, NY	293,280	6.38
251	Waco, TX	233,378	6.37
252	Jacksonville, NC	173,064	6.36
253	Madera-Chowchilla, CA	148,632	6.36

Notes: Wages from CPS. Averages from tobit regression accounting for top-coding.

Table 3: Hedonic regressions for rental and owner-occupied units.

	log rent		log value	
Number of rooms				
2	0.1344***	(0.0018)	0.1753***	(0.0058)
3	0.1793***	(0.0019)	0.3770***	(0.0055)
4	0.2703***	(0.0019)	0.3914***	(0.0055)
5	0.3345***	(0.0020)	0.5437***	(0.0055)
6	0.4182***	(0.0021)	0.7087***	(0.0055)
7	0.4933***	(0.0025)	0.8805***	(0.0055)
8	0.5470***	(0.0029)	1.0411***	(0.0055)
9+	0.5839***	(0.0032)	1.3040***	(0.0055)
Age of structure				
2-5 years	-0.0332***	(0.0034)	-0.0516***	(0.0014)
6-10 years	-0.0978***	(0.0033)	-0.1260***	(0.0014)
11-20 years	-0.1836***	(0.0031)	-0.2441***	(0.0013)
21-30 years	-0.2612***	(0.0031)	-0.3692***	(0.0013)
31-40 years	-0.3145***	(0.0031)	-0.4310***	(0.0014)
41-50 years	-0.3560***	(0.0032)	-0.4818***	(0.0014)
51-60 years	-0.3974***	(0.0032)	-0.5579***	(0.0014)
61+ years	-0.3772***	(0.0031)	-0.5606***	(0.0014)
Units in structure				
1-family detached	0.0366***	(0.0028)	0.0058***	(0.0010)
1-family attached	-0.0439***	(0.0030)	-0.2001***	(0.0013)
2-family	-0.0509***	(0.0030)	0.0201***	(0.0018)
3-4 family	-0.0651***	(0.0029)	0.0216***	(0.0023)
5-9 family	-0.0702***	(0.0030)	-0.1520***	(0.0029)
10-19 family	-0.0184***	(0.0030)	-0.1924***	(0.0034)
20-49 family	-0.0499***	(0.0031)	-0.0745***	(0.0033)
50+ family	-0.0281***	(0.0030)	0.0510***	(0.0026)
Dummy kitchen	0.0417***	(0.0031)	0.3138***	(0.0035)
Constant	6.0817***	(0.0060)	10.8201***	(0.0067)
City Effects	yes		yes	
Within-R2	0.0675		0.283	
N	3,274,198		7,680,728	
Cities	533		533	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

!! Reference group for units in structure is N/A and needs to be changed.

Table 4: Rank of cities by estimated housing price index.

	City	Population	Rent Index
1	Honolulu, HI	907,574	1312.76
2	San Jose-Sunnyvale-Santa Clara, CA	1,839,700	1252.62
3	Los Angeles-Long Beach-Santa Ana, CA	12,874,797	1157.05
4	San Francisco-Oakland-Fremont, CA	4,317,853	1148.75
5	Washington-Arlington-Alex., DC-VA-MD-WV	5,476,241	1142.28
6	New York-New Jersey-Long Isl., NY-NJ-PA	19,069,796	1118.46
7	Santa Barbara-Santa Maria-Goleta, CA	407,057	1117.70
8	Oxnard-Thousand Oaks-Ventura, CA	802,983	1095.97
9	Santa Cruz-Watsonville, CA	256,218	1012.10
10	San Diego-Carlsbad-San Marcos, CA	3,053,793	1001.02
244	Lawton, OK	113,228	330.28
245	Anniston-Oxford, AL	114,081	329.33
246	Saginaw-Saginaw Township North, MI	200,050	325.06
247	Huntington-Ashland, WV-KY-OH	285,624	323.11
248	Decatur, AL	151,399	321.53
249	Brownsville-Harlingen, TX	396,371	320.61
250	Flint, MI	424,043	316.47
251	Johnstown, PA	143,998	307.73
252	Monroe, LA	174,086	297.80
253	McAllen-Edinburg-Mission, TX	741,152	291.55

Notes: Housing price indices based on hedonic regressions using 2000 U.S. Census data and adjusted for 2000-2009 price changes with reapeat-sales indices from the Federal Housing Finance Agency.

Table 5: Rank of cities by average of log utility.

	City	Population	Average utility
1	Ann Arbor, MI	347,563	4.91
2	Bridgeport-Stamford-Norwalk, CT	901,208	4.89
3	Jackson, MI	159,828	4.88
4	Barnstable Town, MA	221,151	4.87
5	Flint, MI	424,043	4.87
6	Worcester, MA	803,701	4.85
7	Detroit-Warren-Livonia, MI	4,403,437	4.82
8	Boulder, CO	303,482	4.82
9	Decatur, AL	151,399	4.81
10	Hartford-West Hartford-East Hartford, CT	1,195,998	4.81
244	Corpus Christi, TX	416,095	4.37
245	Honolulu, HI	907,574	4.36
246	Laredo, TX	241,438	4.36
247	El Paso, TX	751,296	4.34
248	Lynchburg, VA	247,447	4.32
249	Utica-Rome, NY	293,280	4.32
250	Chico, CA	220,577	4.30
251	Madera-Chowchilla, CA	148,632	4.28
252	Jacksonville, NC	173,064	4.28
253	Amarillo, TX	246,474	4.25

Table 6: Rank of cities by variance of log utility.

	City	Population	S.D. Utility
1	Santa Cruz-Watsonville, CA	256,218	0.75
2	Punta Gorda, FL	156,952	0.74
3	Boulder, CO	303,482	0.70
4	Springfield, OH	139,671	0.69
5	Springfield, IL	208,182	0.68
6	San Jose-Sunnyvale-Santa Clara, CA	1,839,700	0.67
7	Durham-Chapel Hill, NC	501,228	0.66
8	Bridgeport-Stamford-Norwalk, CT	901,208	0.66
9	San Francisco-Oakland-Fremont, CA	4,317,853	0.65
10	Lubbock, TX	276,659	0.65
244	South Bend-Mishawaka, IN-MI	317,538	0.42
245	Myrtle Beach-North Myrtle Beach-Conway, SC	263,868	0.41
246	Anderson, IN	131,417	0.41
247	Warner Robins, GA	135,715	0.41
248	Lynchburg, VA	247,447	0.40
249	Appleton, WI	221,894	0.40
250	Janesville, WI	160,155	0.37
251	Waco, TX	233,378	0.37
252	Macon, GA	231,576	0.36
253	Panama City-Lynn Haven, FL	164,767	0.35

References

[1]

[2] Albouy, David "Are Big Cities Really Bad Places to Live? Improving Quality-Of-Life Estimates Across Cities", NBER Working Paper 14472, 2008.

[3] Baum-Snow, Nathaniel, and Ronni Pavan, "Understanding the City Size Wage Gap," 2010a mimeo.

[4] Baum-Snow, Nathaniel, and Ronni Pavan, "Inequality and City Size", mimeo, 2010b.

[5] Behrens, Kristian, Gilles Duranton and Frederic Robert-Nicoud "Productive cities: Sorting, Selection and Agglomeration", mimeo, 2010.

[6] Case, K. E, and R. J. Shiller, "A Multivariate Repeat-Sales Model for. Estimating House Price Indices", *Journal of Urban Economics* 57, 1987, 320-342.

[7] Ciccone, Antonio, and Giovanni Peri, "Long-Run Substitutability Between More and Less Educated Workers: Evidence From U.S. States, 1950–1990," *Review of Economics and Statistics,* **87(4)**, 2005, 652-663

[8] Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sebastien Roux "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection", mimeo, 2009.

[9] Davis, Morris A., Jonas D.M. Fisher, and Toni M. Whited, "Macroeconomic Implications of Agglomeration," U. Wisconsin mimeo, 2010.

[10] Davis, Morris A., and Francois Ortalo-Magné, "Household Expenditures, Wages, Rents," *Review of Economic Dynamics* 2009.

[11] Eeckhout, Jan, "Gibrat's Law for (all) Cities", *American Economic Review* **94(5)**, 2004, 1429-1451.

[12] Eeckhout, Jan, and Philipp Kircher, "Sorting and Decentralized Price Competition," *Econometrica* **78(2)**, 2010, 539-574.

[13] Eeckhout, Jan, and Roberto Pinheiro, "Diverse Organizations and the Competition for Talent," mimeo 2010.

[14] Gabaix, Xavier. "Zipf's Law for Cities: An Explanation", *Quarterly Journal of Economics* **114(3)**, 1999, 738-767.

[15] Gautier, Pieter, M. Svarer and C.N. Teulings,, "Marriage and the City: Search Frictions and Sorting of Singles," *Journal of Urban Economics* **67(2)**, 2010, 206-218.

[16] Gould, Eric, "Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium," *Review of Economic Studies* **74**, 2007, 477-506.

[17] Helpman, Elhanan, Oleg Itskhoki, and Stephen Redding "Inequality and Unemployment in a Global Economy," *Econometrica*, forthcoming, 2010.

[18] Holmes, Thomas J., and John J. Stevens, "Geographic Concentration and Establishment Size: Analysis in an Alternative Economic Geography Model," University of Minneapolis mimeo, 2003.?

[19] Keane, Michael, and Kenneth Wolpin, "The Career Decisions of Young Men," *Journal of Political Economy* **105(3)**, 1997, 473-522.

[20] Lucas, Robert E. Jr., and Esteban Rossi-Hansberg, "The Internal Structure of Cities", *Econometrica* **70(4)**, 2002, 1445-1476.

[21] Moretti, Enrico "Real Wage Inequality", mimeo, 2010.

[22] Munshi, Kaivan, "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market," *Quarterly Journal of Economics* **118(2)**, 2003, 549-597.

[23] Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek, "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]", Minneapolis: University of Minnesota, 2010.

[24] Sieg, Holger, V. Kerry Smith, H. Spencer Banzhaf, Randy Walsh, "Interjurisdictional housing prices in locational equilibrium", *Journal of Urban Economics*, 52, 2002, 131-153.