CESifo Working Papers
www.cesifo.org/wp

Spatial Sorting:
Why New York, Los Angeles and Detroit Attract the Greatest Minds as well as the Unskilled

Jan Eeckhout
Roberto Pinheiro
Kurt Schmidheiny

CESifo Working Paper No. 3274
Category 4: Labour Markets
December 2010

CESifo *Working Paper No. 3274*

Spatial Sorting:
Why New York, Los Angeles and Detroit Attract the Greatest Minds as well as the Unskilled

Abstract

We propose a theory of skill mobility across cities. It predicts the well documented city size-wage premium: the wage distribution in large cities first-order stochastically dominates that in small cities. Yet, because this premium is reflected in higher house prices, this does not necessarily imply that this stochastic dominance relation also exists in the distribution of skills. Instead, we find 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. The pattern of spatial sorting is explained by a technology with a varying elasticity of substitution that is decreasing in skill density. Using CPS data on wages and Census data on house prices, we find that this technology is consistent with the observed patterns of skills.

JEL-Code: J31, R10, R23.

Keywords: matching theory, sorting, general equilibrium, population dynamics, cities, wage distribution.

Jan Eeckhout
ICREA-GSE-UPF and
University of Pennsylvania
Philadelphia / USA
jan.eeckhout@gmail.com

Roberto Pinheiro
Leeds School of Business
University of Colorado
Boulder / USA
Roberto.Pinheiro@colorado.edu

Kurt Schmidheiny
Department of Economics
University Pompeu Fabra
Barcelona / Spain
kurt.schmidtheiny@upf.edu

November, 2010
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.

"*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.

In this paper we show that there is indeed evidence of spatial sorting and that disproportionately more skilled citizens locate in larger cities. However, we provide a key new insight: larger cities also disproportionately attract lower skilled agents. For example, in New York city there is a huge low skill contingent in the South Bronx and Newark as well as the high skilled mainly living in Manhattan. Similarly, while Detroit has disproportionately many low skilled individuals and a reputation for inner city poverty, it also disproportionately attracts high skilled individuals, many of whom live in the wealthy neighborhood of Bloomfield Hills. In that respect, large cities like New York and Detroit are more similar to each other than to small cities. We show that there is a systematic pattern of fat tails in the skill distribution of large cities. To our knowledge, this pattern of spatial sorting has not been documented in the literature.

We propose a theory of city choice and heterogeneous skills that rationalizes this pattern of fat tails for larger cities. The main innovation of our theory is that it generates a pattern of sorting that does not involve perfect segregation of skills across cities. Consistent with reality, in our model cities attract *all* skill types, yet to a varying extent. Citizens earn a living based on a competitive wage, and under perfect mobility, their location choice will make them indifferent between consumption-housing bundles, and therefore between different wage-house price pairs across cities. 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).[1] Output is produced with heterogeneous labor inputs. The technology 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. Since under the assumptions of our model, in equilibrium there is a one-to-one relationship between wages and skills within each city, we are able to use revealed preference choices of location and wages

[1] 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. 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.

paid to back out skills.

We find that, for general production technologies, the size of the city is 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 skill distribution is crucially determined by the technology. In the benchmark case of Constant Elasticity of Substitution (CES), cities of different TFP have different population sizes, but the distribution of skills is the same across cities, and for that matter across the entire economy. In contrast, when the elasticity of substitution varies across skill levels, distributions across cities differ. In particular, when the elasticity is decreasing in the measure of a given skill, then larger cities have skill distributions with fatter tails.

Our empirical analysis documents the systematic sorting pattern that leads to fat tails in large cities. From the theory we construct a price-theoretic measure of skills, based on revealed preference location choices given wages and house prices. Consider first the distribution of wages. 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.

However, 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. From the theory, the resulting distribution of utilities is therefore 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. Our main finding is that 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. This systematic pattern of spatial sorting is extremely robust as we document in the discussion: we consider different definitions of large vs. small cities, use three different data sources for local housing values, include local price differences in consumption goods and analyze the observable part of skills by educational attainment.

The fat tails imply that the relative demand for skills is non-monotonic. In large cities, there is relatively high demand for low and high skilled workers, whereas the demand for medium skilled workers

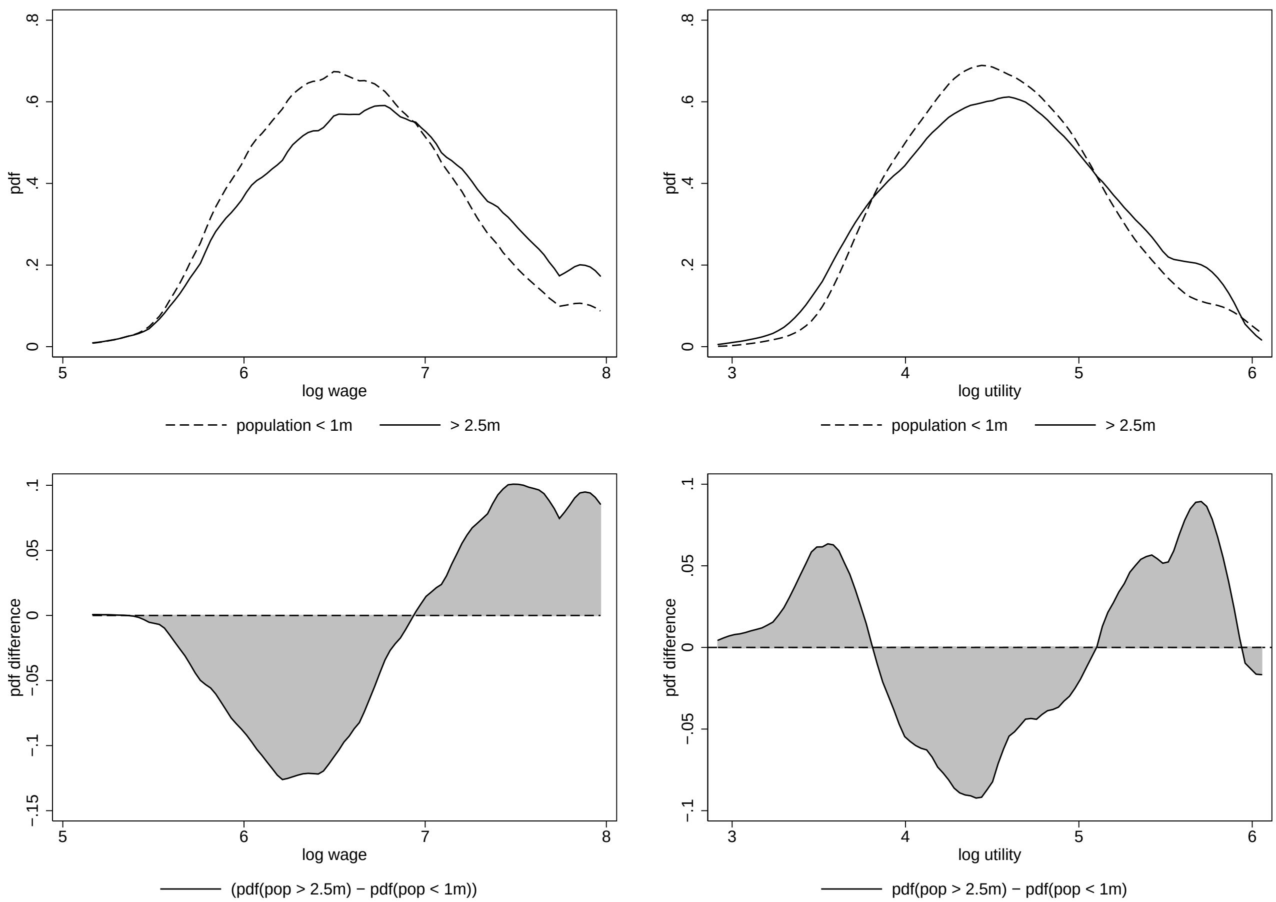

Figure 1: Left-to-right-top-to-bottom. A. Wage distribution for small and large cities; B. Skill distribution for small and large cities; C. Density differential of wages between large and small cities; D. Density differential of skills between large and small cities.

is relatively low. Figure 1.D illustrates this non-monotonic relative demand pattern. In contrast, relative wages are monotonic (Figure 1.C).

A key feature of our approach is the price-theoretic measure of skills. This is in contrast to the common approach of using observable skills such as years of education or test scores. Not only do observables explain a mere fraction of skills (see for example Keane and Wolpin (1997)), observed skill categories are typically very coarse. In the literature, skills are often partitioned into two classes,[2] which allows for inference of a linear approximation when the underlying relation is monotonic. 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 – there is no hope to satisfactorily identify this non-monotonic relation with two points only. Once we allow for many skill classes using the wage based measure, we are able to characterize a smooth distributions of skills.

[2]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.

In addition to our wage based measure, we nonetheless also analyze the skill distribution based on observable skills, 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.

2 Related Literature

There is a long tradition in the Urban Economics literature investigating differences across city sizes, in particular with respect to varying standards of living between cities. We are also 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.

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 predicted values as an index for local cost-of-living differences. Differences in real wages across MSAs are interpreted as quality-of-life 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.

Our findings are also related to the previous literature on variations of skill distributions across city sizes. Bacolod, Blum and Strange (2009) study the difference in skill distributions across city sizes, using jointly Census and NLSY data and the Dictionary of Occupational Titles (DOT), defining skills as a combination of qualities instead of just education. They find a small variation in cognitive, people, and motor skills across city sizes, which they attribute to skills being defined nationally, not being able to address local differences in occupational requirements of skills. Once they look at differences in the Armed Forces Qualification Test (AFQT) and the Rotter Index – measures of intelligence and social skills, respectively – they find that, even though the average scores are quite similar across city sizes, the scores at large cities for the lowest scores (10th percentile) are much lower than the ones at small cities. Similarly, the highest scores (90th percentile) were much higher in large cities than small ones. These results corroborate the idea that we have fat tails in the distribution of skills, even though differences in average skill may be small.

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. The work by Behrens, Duranton and Robert-Nicoud (2010) also analyzes sorting of heterogeneous agents into cities. They find that more productive workers locate in large cities and less productive workers in small cities. As a result, they predict as we do the effect in the upper tail, however not that in the lower tail.

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 and given by the order: $i \in \mathcal{I} = \{1, ..., I\}$. Associated with this skill order is a level of productivity $y_i^\beta, \beta > 0$, where 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.

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, etc. 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

[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.

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 representing a city's productive amenities, infrastructure, historical industries, etc.[4] In each city, firms compete to operate in this market. Firms are 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 k_j that in general is city specific. In particular, we will assume that firms need to buy a constant amount of k units of land in the city in order to engage in economic activity. Firms need to rent housing space for production and house prices affect the expenditure that a firm incurs to finance infrastructure. As a result, the firm's entry cost is kp_j.[5] Competitive entry will drive down profits to zero, which are given by:

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

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 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] 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,...

[5] This particular assumption is not crucial for any of our results. We make the assumption because it is realistic and in some of the derivations the dependence on p_j simplifies the expressions.

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

$$N_j\left(k+\sum_{i=1}^{I}h_{ij}m_{ij}\right)=H_j,\ \forall j.$$

Within a city there are a measure of N_j identical firms, each of which demands k units of land to operate in the market and each employs m_{ij} skilled workers of each skill i who demand h_{ij} units of land for housing.

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)

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

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:

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

This allows us to link the wage distribution across different cities j,j'. Wages across cities relate as:

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

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

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

The first-order condition is:[6]

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

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). It is conceivable that the second effect dominates the first effect in the upward sloping part of the skill density. Suppose skills barely increase in i, yet the density is very steep. Adjacent skills are very abundant even though they are more productive. As a result, the wages of the higher skilled types could be lower. Instead, for a given skill distribution, if productivity is sufficiently increasing, i.e. dy_i^{β}/di is sufficiently large, then wages will always be increasing. In order to avoid the possible reordering of skills, we assume that wages are monotonic in the original order i by making the following mild assumption.

Assumption 1 *Productivity-skill monotonicity. For a given economy, there exists a critical $\beta^{\star}$ such that for every $\beta > \beta^{\star}$ productivity is increasing in skill i in every city j.*

Since wages are equal to marginal product, this implies that wages are increasing in the order of skills i. This is without loss of generality if one interprets the order of skills to be determined by the marginal product, i.e., we have a price-theoretic foundation for skill. Of course, even if this assumption is not satisfied, the analysis still goes through. We can reorder skill types and replace i by an order $\tilde{i}$ such that wages are increasing. Then the analysis applies for the distribution of skills on the order $\tilde{i}$. 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. In our price theoretic view of skills, the account would be more skilled than the artist.

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

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

condition in both cities we can substitute for m_{ij} to obtain for any two cities j, j':

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

Then we can write the wage ratio as:

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

Perfect mobility of consumers equalizes utility and implies $\frac{w_{i1}}{w_{i2}} = \left(\frac{p_1}{p_2}\right)^\alpha$. Hence, 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 analogously for w_{i2}.

Even though we have not closed the model yet, it is important at this stage to note that wages depend only on the quantity of aggregate skills M_i and not on the city level quantity m_{ij}. This underscores the importance of the general equilibrium effect from mobility: quantities respond to arbitrage until the wage-price ratio is constant across all pairs, and as a result, wages are pinned down only by the economy-wide supply of skills. Naively taking the first order condition $w_{ij} = \gamma_i A_j (m_{ij})^{\gamma_i-1} y_i^\beta$ to the data by regressing wages on the quantity of labor in a given city is therefore completely uninformative, except in the unrealistic case of zero mobility.

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. We can therefore fully characterize the equilibrium by the following four equations in four unknowns p_1, p_2, N_1, N_2:

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

$$\sum_{i=1}^{I}\left(\frac{1}{1+\frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}}}\frac{M_i}{N_2}\right)^{\gamma_i}\left(1-\left(1-\alpha\right)\gamma_i\right)y_i^{\beta} = \frac{p_2}{A_2}\frac{H_2}{N_2} \tag{2}$$

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

$$\sum_{i=1}^{N}\left(\frac{1}{1+\frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}}}\frac{M_i}{N_2}\right)^{\gamma_i}\left(1-\gamma_i\right)y_i^{\beta} = \frac{k}{A_2}p_2 \tag{4}$$

In what follows we refer to these as the four equilibrium conditions (1)–(4).

The Main Theoretical Results. First we consider the 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 in a city with an extremely tiny geographical area, labor demand 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 2 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.

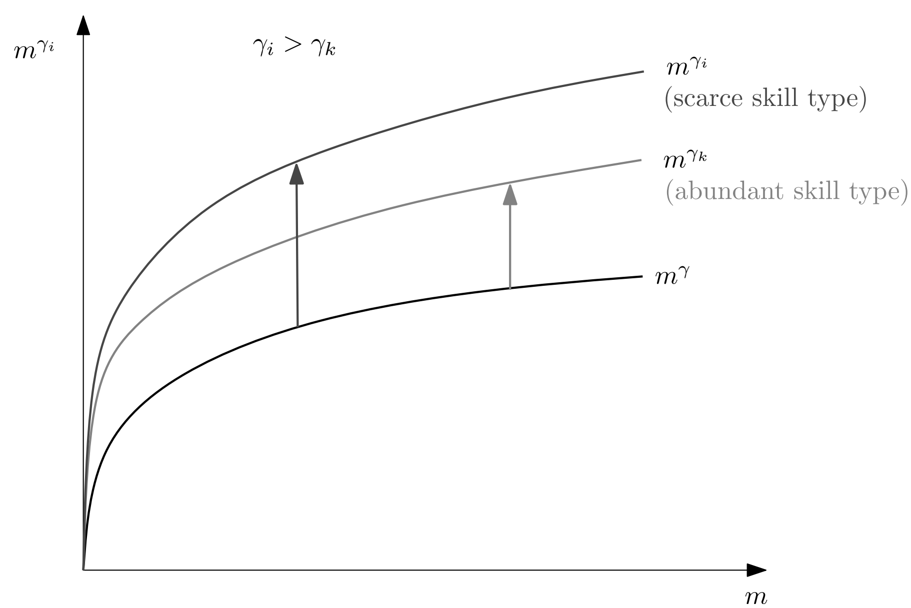

Figure 2: Scarce skill types have a higher γ_i, implying a higher level of productivity *and* a higher a marginal productivity.

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. Moreover, in recent work Fu and Ross (2010) find little evidence of sorting within metropolitan areas based on agglomeration.

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

Proof. In Appendix. ■

To see the intuition behind this result, consider first 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. Now consider the case of VES, by increasing the marginal product for low and high skilled workers, leaving that of the medium skilled at the CES level. This increase in marginal productivity will be larger in large cities because they have higher TFP and TFP and skills are complementary. As a result, in large cities low and high skills will experience a higher increase in productivity and therefore in wages relative to medium skills, and vice versa in small cities. This cannot be offset by higher house prices because those are determined by real wage equalization at all skill levels, including the medium skilled. The higher real wages for low and high skilled workers in large cities will attract those skill types into the large cities driving down nominal wages until real wages equalized. This in-migration of low and high skilled workers leads to the fat tails in the large cities.

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* $p_j h^\star_{ij}$ *is higher in larger cities. The size of houses* $h^\star_{ij}$ *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 $h^\star_{ij}$ 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. It immediately follows from the proof of Proposition 1 that firm size is increasing in city size. 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 representative firm is larger in larger cities. Firm size is given by $S_j = N_j \sum_i m_{ij}$, and is increasing in city size. It is ambiguous whether the number of firms N_j is larger in larger cities.

Corollary 2 *Firms are larger in the larger cities.*

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 in fixed proportions 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.

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

$$U_i = \alpha^{\alpha} (1-\alpha)^{1-\alpha} \frac{w_{ij}}{p_j^{\alpha}} \tag{5}$$

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.

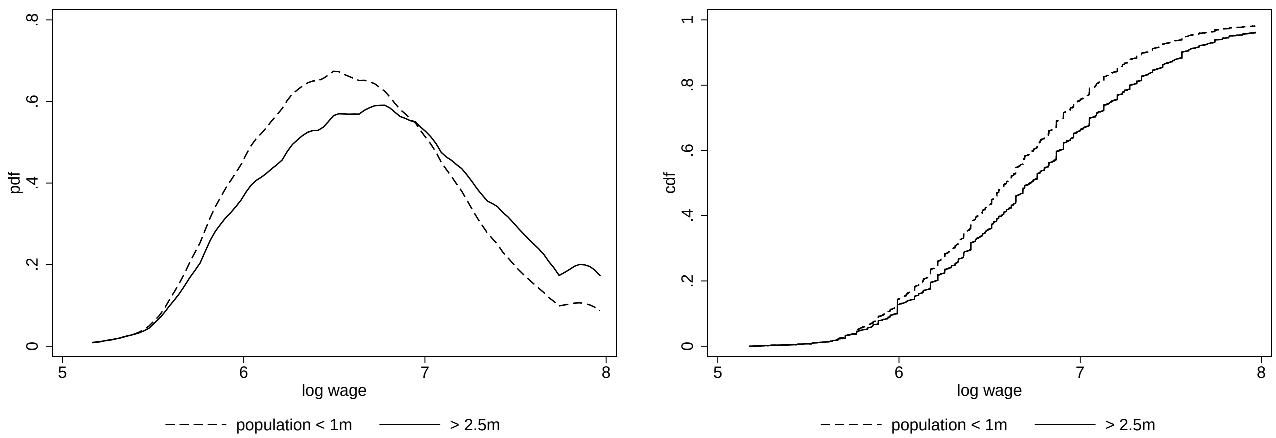

Figure 3: Wage distribution for small and large cities. Full-time wage earners from 2009 CPS. A. Kernel density estimtates (Epanechnikov kernel, bandwidth = 0.1), *not* adjusted for top-coding; B. Empirical CDF, accounting for top-coding.

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 the 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

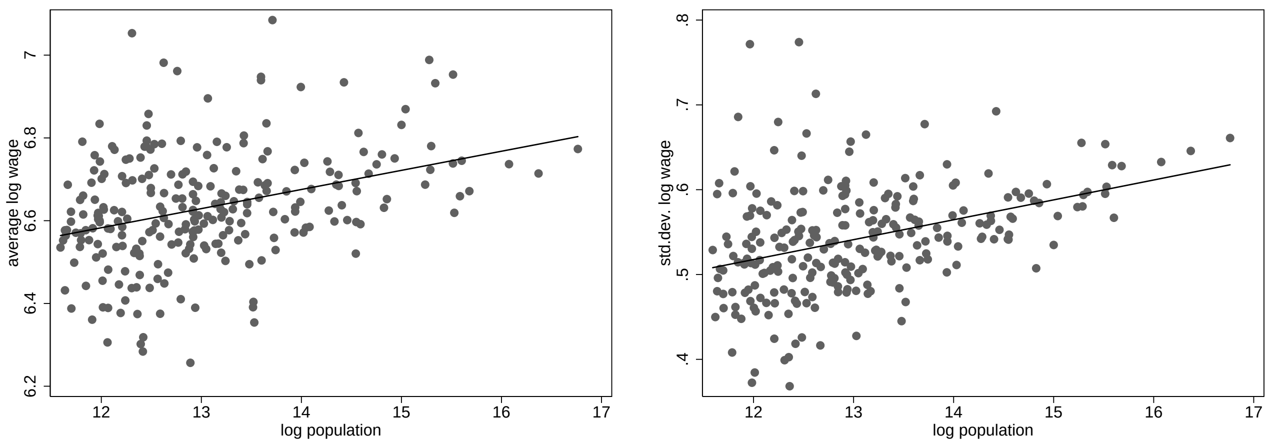

Figure 4: Wage distribution by population size. A. Mean (slope=0.046, s.e.=0.007); B. Standard Deviation (slope=0.023, s.e.=0.003). Based on censored regression accounting for top-coding.

in large cities.[8] A simple t-test shows that wages in large cities are 13.2% higher than in small ones ($t = 28.3, p < 0.000$). Controlling for right censoring from top-coding and weights in a censored (tobit) regression leads to almost exactly the same comparison: Δ log wage= 13.1% (robust $t = 24.7$, $p < 0.000$). Figure 16 in the appendix draws the confidence intervals of the distribution and also shows that the difference is significant.

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 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 $t = 6.89$, $p > 0.000$) and of 0.023 (robust $t = 7.87$, $p > 0.000$) 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

[8]Note that the "bumps" in the top tail for both large and small cities are an artefact of the top-coded *nominal* wage data.

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 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 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.

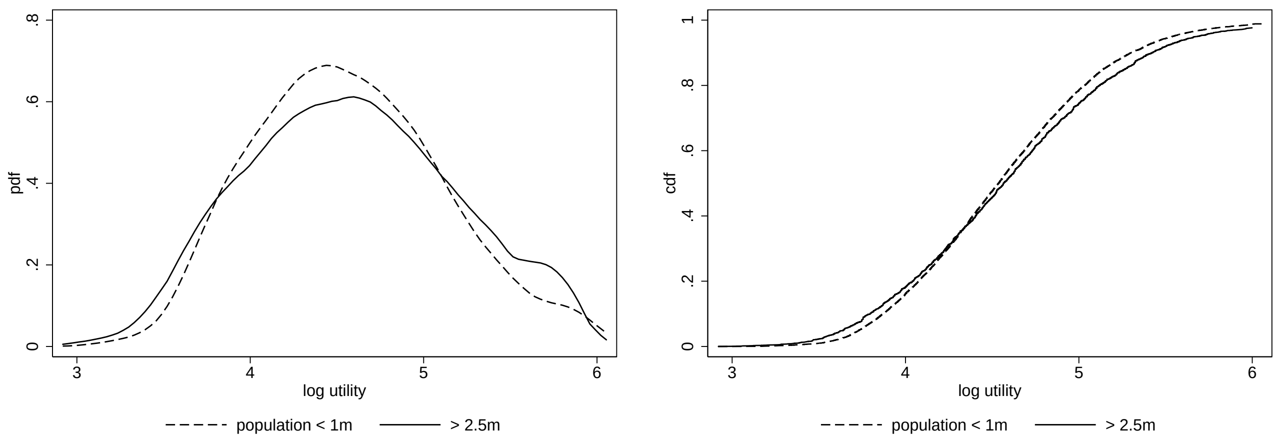

Figure 5: Skill distribution for small and large cities. A. Kernel density estimates (Epanechnikov kernel, bandwidth = 0.1), *not* adjusted for city-specific top-coding; B. Empirical CDF adjusted for top-coding using the Kaplan-Meier method.

[Insert Table 4 here]

5.5 Skill distribution

Davis and Ortalo-Magné (2007) document that expenditure shares on housing are remarkably 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 (5).

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.[9] Figure 16 in the appendix also draws confidence intervals of the distribution and shows that the difference is significant.

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. 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 0.016 for mean log utility (robust $t = 2.21$, $p > 0.028$)

[9] Note again that the "bumps" in the top tail are due to top coding, see footnote 8. Top-codes appear more to the left for large cities because *real* wages are deflated with higher house prices.

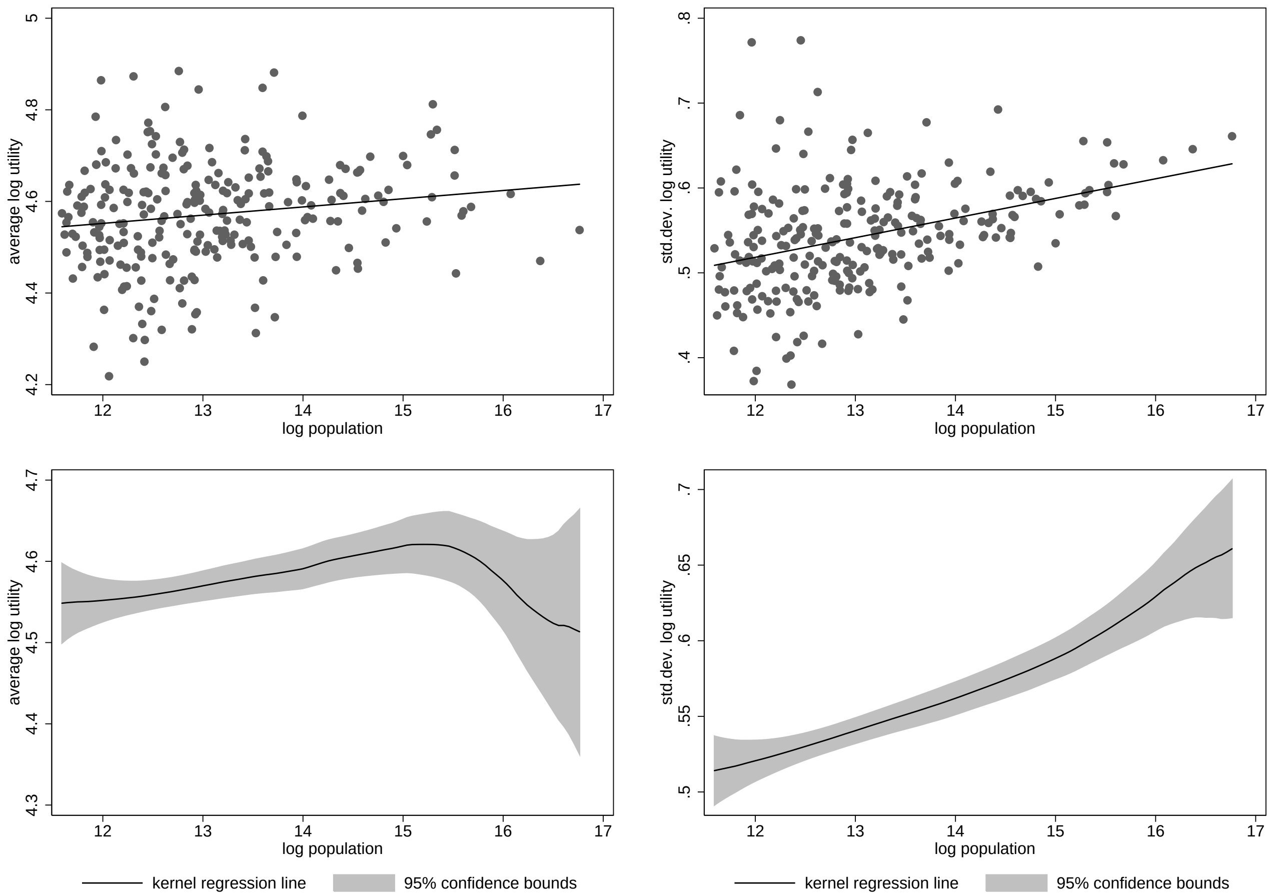

Figure 6: Skill distribution by population size. Left graphs: Mean; Right graphs: Standard Deviation. Top graphs: linear regression (slope average=0.016 (s.e.=0.007); slope st.dev.=0.023 (s.e.=0.003)). Bottom graphs: local linear regression (Epanechnikov kernel, bandwidth = 0.58 and 0.74, respectively)

and of 0.023 (robust $t = 7.61$, $p > 0.000$) 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]

As we did in the introduction, to emphasize the fat tails result, in Figure 7 we reproduce the difference of the density functions. In panel A, the ratio of wages in large cities relative to small cities is increasing in wages, thus illustrating that relative wages are monotoic. In panel B, the relative densities of the utilities is non-monotonic. It is important here to point out that the decrease in the pdf differences at the very top skill levels is driven by the top coding. Note that for large cities the impact of top coding is further to the left due to on average higher house prices.

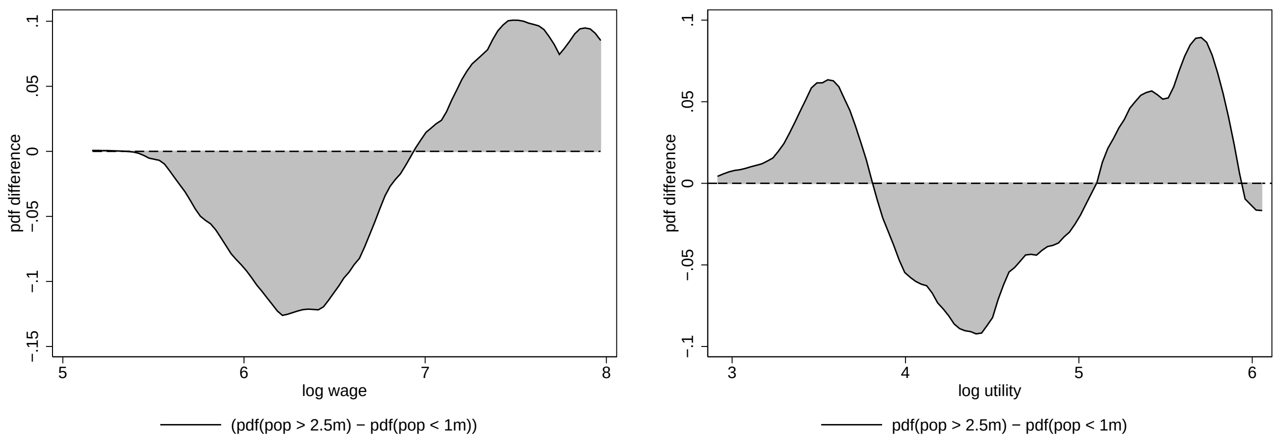

Figure 7: A. Density differential of wages between large and small cities; B. Density differential of skills between large and small cities.

6 Direct Measures of Skills

As a robustness check and as external validation, we compare our implicit skill distribution with observed measures of skill. Figure 8.A. shows the distribution of educational attainment for the same CPS population as our wage data, where workers are grouped in 7 education categories. 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.

This can be observed even more transparently when we group the education levels into three groups. This is reported in Figure 8.B. What is most striking about this observation is that the fat tails in the distribution of educational attainment is obtained *independently* of how we constructed our measure of skills before. Here, 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.

Using observable, self-reported measures of skills – either education categories or years of schooling – we find a distribution with fatter tails in larger cities, both in the aggregate and at the individual city level.

The key identifying assumption to derive the skill distribution from wages is perfect mobility of identically skilled workers. For our wage-based skill measure that implies that utility of a given skilled

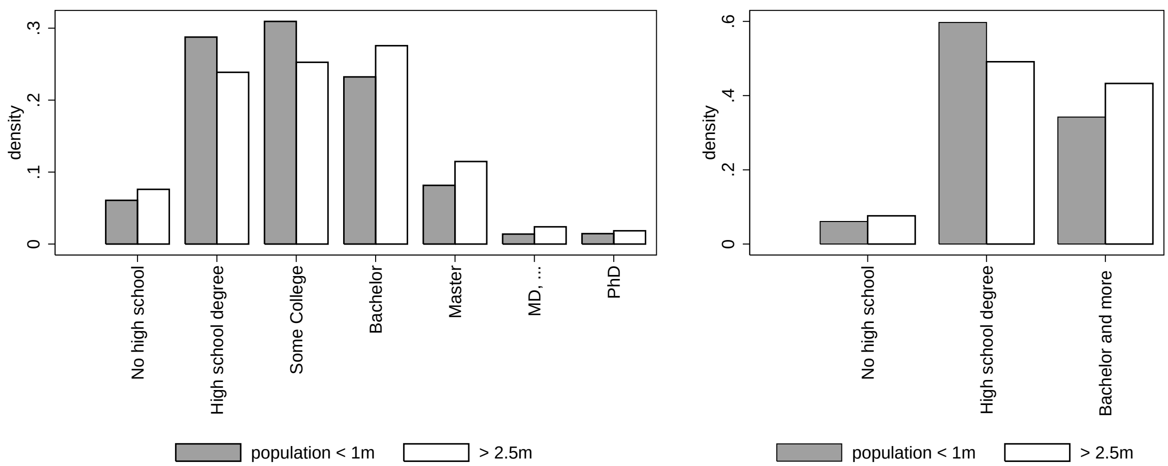

Figure 8: Observed educational attainment for small and large cities. Highest completed grade of full-time wage earner in 2009 CPS. A. Grouped in 7 categories; B. Grouped in 3 categories.

worker is the same across different cities. Here we can verify whether this assumption holds when we use *observable* skills instead. Note that utility need not be equalized for identical observed skill levels across different cities when our predicted skill level is imperfectly correlated with the observable skill measure. Nonetheless, it is instructive to investigate how average utility (wages corrected for house prices) vary across different city size conditional on the observed skill group.

In Table 7 we report the linear regression by observable skill group of the average utility and of its standard deviation on city size (Figure 17 depicts for each skill category the scatter plot for each city together with the regression line). Before discussing the findings, an important caveat is due. By dividing workers in subgroups, some of the subgroups include city-education subgroup that have not enough observations to calculate the mean and standard deviation. Those with city-education subgroups with less than two observations are dropped. The lack of observations is most acute in the highest skill categories. Table 7 reports the number of cities N in each skill category out of all 253 cities where there are at least two observations. Because the censoring of the data may well be systematic, these results should be taken as merely indicative.

[Insert Table 7 here]

For what the data is worth, we find that by observable education category, in 6 out of the 7 groups utilities (real wages) do not significantly vary with city size. The one exception is the group with the Master degrees. This effect is strong enough to render the overall effect to be positive as well, though the effect is small. This seems to indicate that there is some systematic variation of wages across cities for this education group. For example, there could be systematic variation in the location decision and the enrollment in masters degrees, which predominantly happens *after* several years of work experience (e.g., MBA degrees).

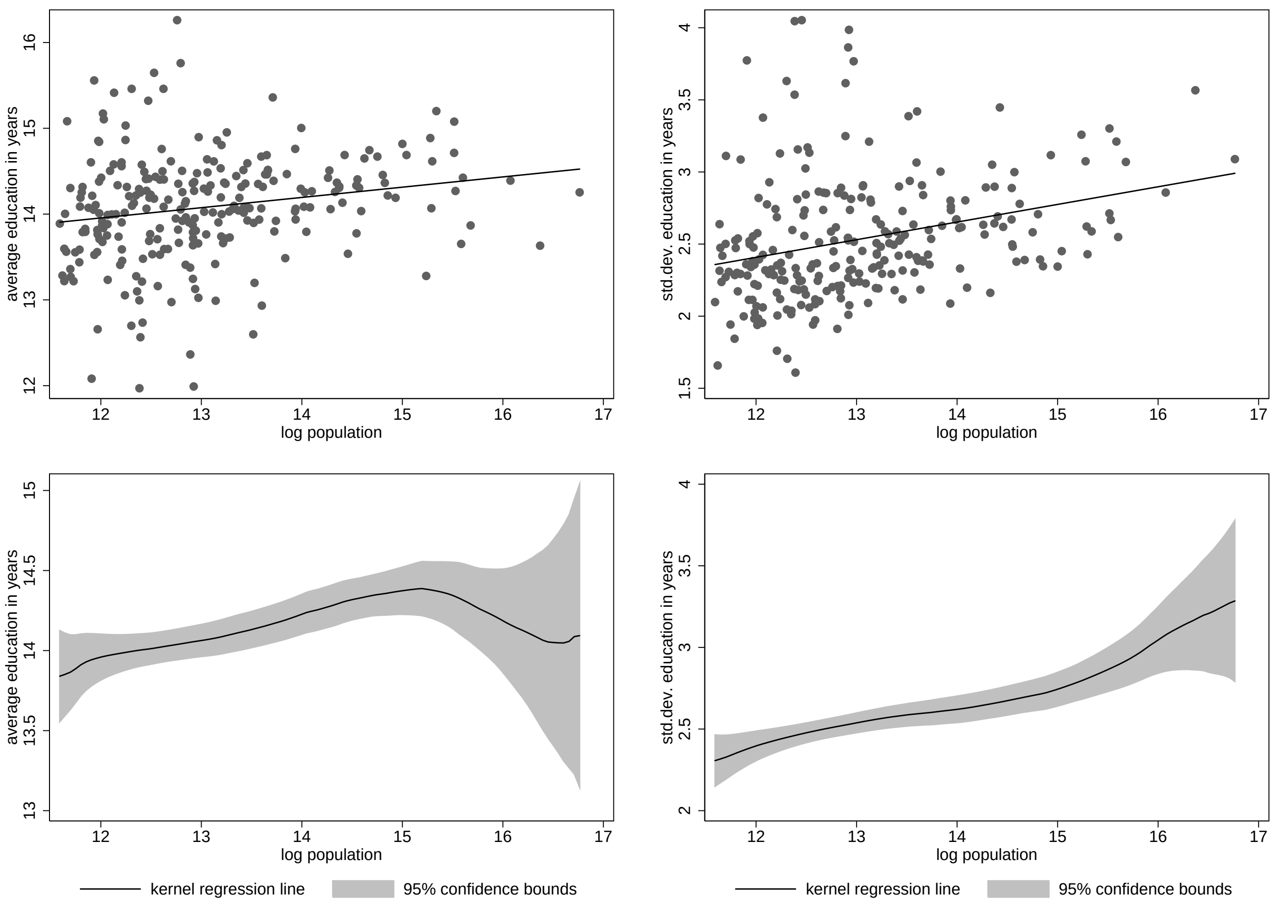

Figure 9: Distribution of educational attainment (translated into years) by population size. Left graphs: Mean; Right graphs: Standard Deviation. Top graphs: linear regression (slope average=0.12 (s.e.=0.03); slope st.dev.=0.12 (s.e.=0.02)). Bottom graphs: local linear regression (Epanechnikov kernel, bandwidth = 0.53 and 0.67, respectively)

This indicates that on average our price-theory measure of skills appears to be quite correlated with the measure from observable skills. What the remainder of the table suggests is that even within each observable skill category there is residual heterogeneity in skills. In each skill category, the standard deviation of utility (real wages) is increasing in city size. Even within the observable skill category (degree obtained), there is systematic sorting of the highest and lowest skilled types into large cities and those that are medium skilled locate in medium sized cities. This holds true for *all seven* skill categories. This is consistent with the well-known finding that a large part of wage heterogeneity is not explained by observable skills (see for example Keane and Wolpin (1997)).

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[10] 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 significant. 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 simple inspection of the kernel estimate.[11]

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:

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

where we 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 to 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 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

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

[11]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.

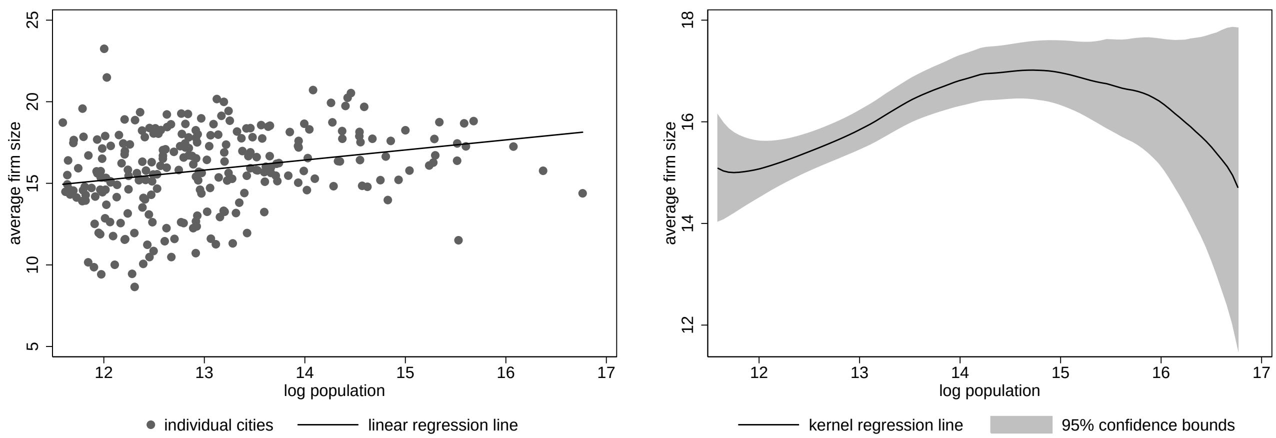

Figure 10: Average firm size by city population: A. linear regression (slope=0.62 (s.e.=0.14)); B. local linear regression (Epanechnikov kernel, bandwidth=0.58).

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.

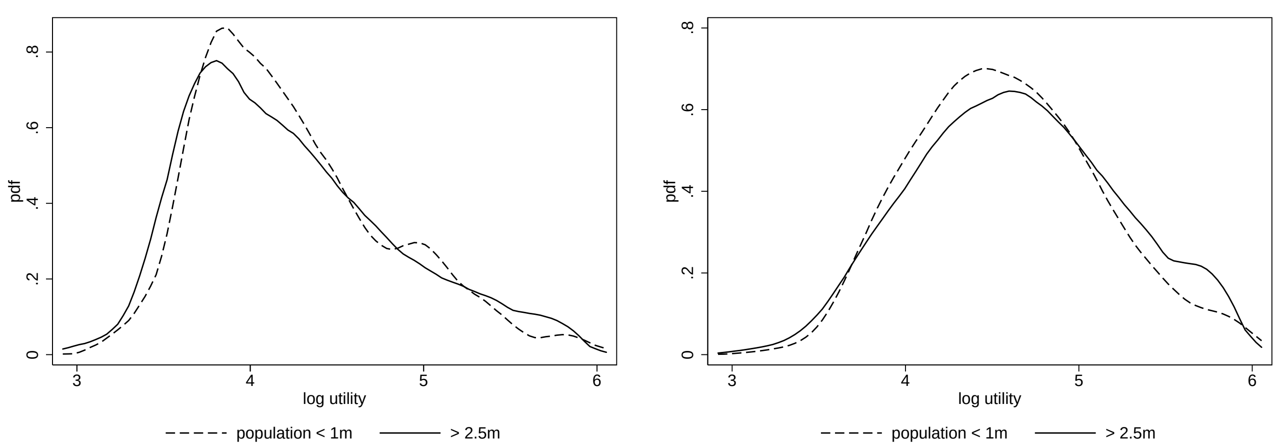

Figure 11: Skill distribution: A. Foreign born workers; B. Natives.

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.

7.3 Alternative Measures of Local Housing Prices

Local housing prices are crucial in our strategy to back out skills from observed nominal wages. In the previous sections, we use rental prices from the 2000 U.S. Census (5% PUMS) and adjust them both for observed hedonic characteristics of the rental units and for the 2000-2009 growth of local house prices. In this section, we check whether our strategy is robust to using different measures of local housing prices.

First, we use house values for owner-occupied units in the 2000 U.S. Census. The advantage over rental prices is that there are about twice as many observed units, that the sub-market of owner-occupied units is more relevant to the majority of households and that the hedonic regression fits the prices data much better (compare the R^2 in Table 3). The big disadvantage is that house values are not market values but the own assessment of the house owner. The top row in figure 12 shows the distribution of skills using house values instead of rental prices. The resulting skill distributions are qualitatively identical to the ones derived from Census rental prices.

Second, we use data for 2009 from the National Association of Realtors (NAR). The house price reflects median sales prices of existing single-family homes by metropolitan statistical area (MSA). This data has the advantage that it reports real market transactions; it also corrects for house characteristics focusing on a single house type only. The disadvantage is that single-family homes are not representative in some metro ares (think of Manhattan) and that the data are contributed from private sources (the realtors) potentially leading to a selective sample with systematic measurement errors. The middle row in figure 12 shows the distribution of skills using house prices from the NAR. Again, the resulting skill distributions are qualitatively identical to the ones derived from Census rental prices.

Third, we use the local housing price index from the ACCRA Cost of Living Index from C2ER (The Council for Community and Economic Research). This price index is a composite from monthly principal and interest payment for a new house (single-family detached house, newly built and not previously occupied) and monthly apartment rents. The advantage of this data is the very explicit combination of owner-occupied and rental units and the very exact description of the representative unit sampled (structure, location, etc.). The disadvantage is again the errors from volunteer data providers (see also the details in the next section). The bottom row in figure 12 shows the distribution of skills using house price index from ACCRA. Once again, the resulting skill distributions are qualitatively identical to the ones derived from Census rental prices.

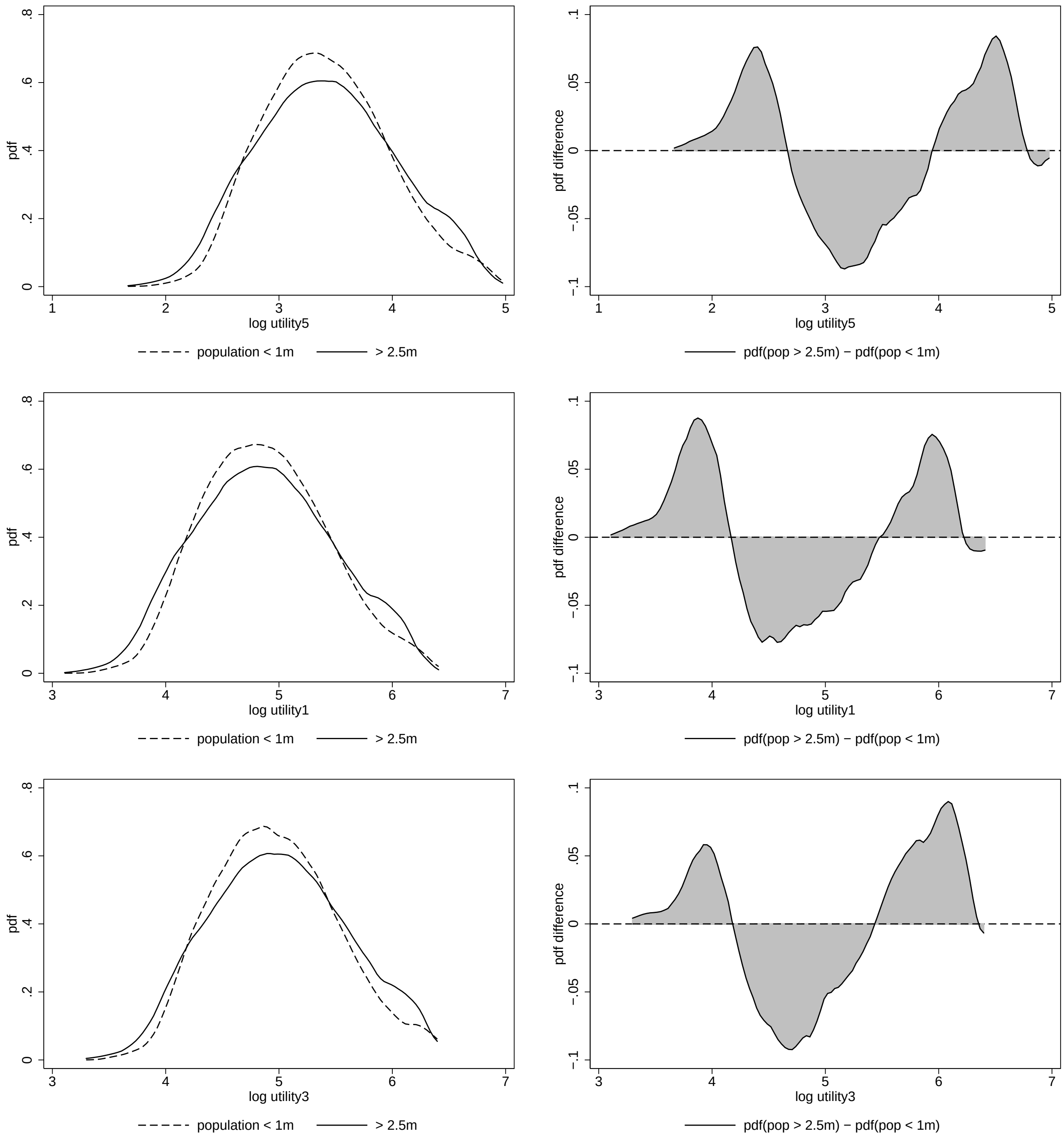

Figure 12: Skill distribution using alternative measures for local housing prices: Top graphs: U.S. Census, house value 2000, adjusted for hedonic characteristics and 2000-2009 price growth; Middle graphs: National Association of Realtors, median house value 2009; Bottom graphs: ACCRA, local housing price index. Left graphs: density; Right graphs: Density differential.

7.4 Variation in Consumption Prices

In this section, we investigate the role of systematic variation in consumption prices across different cities. Maybe consumption prices in large cities are systematically higher than in smaller cities, thus adding also further to the real cost of living in large cities. We use the ACCRA Cost of Living Index from C2ER (The Council for Community and Economic Research). ACCRA reports local prices for 60 goods such as e.g., a sausage, a house, a phone call, gasoline, the drug Lipitor, or a haircut. The data is collected by volunteers from the local chamber of commerce and then used to build price indices for the six broad consumption categories: grocery items, housing, utilities, transportation, health care and services. ACCRA is the only data for price comparisons across a large set of MSAs. Koo et al. (2000) discuss several problems of the ACCRA data. Besides being collected by volunteers and stemming from a very limited set of items, the most fundamental critique is the lack of proper adjustment for quality differences.

The first finding is that the variation in consumption prices is substantially lower than in housing prices (standard deviation across metropolitan areas is 30.1 for the housing prices index compared to 9.6 for grocery items, 14.7 for utilities, 6.7 for transport, 8.9 for health and 6.9 for services; all prices indices are normalized to mean 100).

Figure 13 plots the distribution of skills for large and small cities. The measure is wages adjusted for local price differences in all goods categories reported in the ACCRA data, including housing, consumption goods and services.[12] When including the price index for all consumption and housing, we find that the left tail difference becomes more pronounced while the right tail difference less so. This indicates that consumption prices are systematically higher in larger cities, but to a limited extent since this effect does not annihilate the existing of fat tails. Note again, that the the third crossing at the very top is an artefact of the top-coding (see footnote 9).

These findings should be interpreted with some caution and a few caveats are due. First, the quality of the ACCRA data is dubious. Second, even within a given location, there could be variation in consumption prices paid by skill level. For example, due to different search intensity, the existence of locally segregated markets, etc., the low skilled may end up paying different prices for similar goods within the same city. Using scanner data on household purchases, Broda, Leibtag and Weinstein (2009) find that the poor pay less. Third, data consisting of price indexes and price surveys are likely to not fully account for quality and diversity differences. Due to their size, large cities have more variety on offer and the quality of goods may differ substantially across different cities. Even if a consumer is paying higher prices, a price index incorporating the diversity and quality on offer will be lower. This also appears to be an issue when studying price differences across different countries. Comparing

[12] ACCRA reports a composite price index which is the weighted average of the sixsub-indices, i.e. $P_{composite} = \alpha_{grocery}P_{grocery} + ... + \alpha_{services}P_{services}$, where the αs are the expenditure shares of the six categories summing up to 1. We do not use this aggregation as it is inconsistent with Cobb-Douglas utility. Instead, we use $P_{composite} = (P_{grocery})^{\alpha_{grocery}} \cdot ... \cdot (P_{services})^{\alpha_{services}}$.

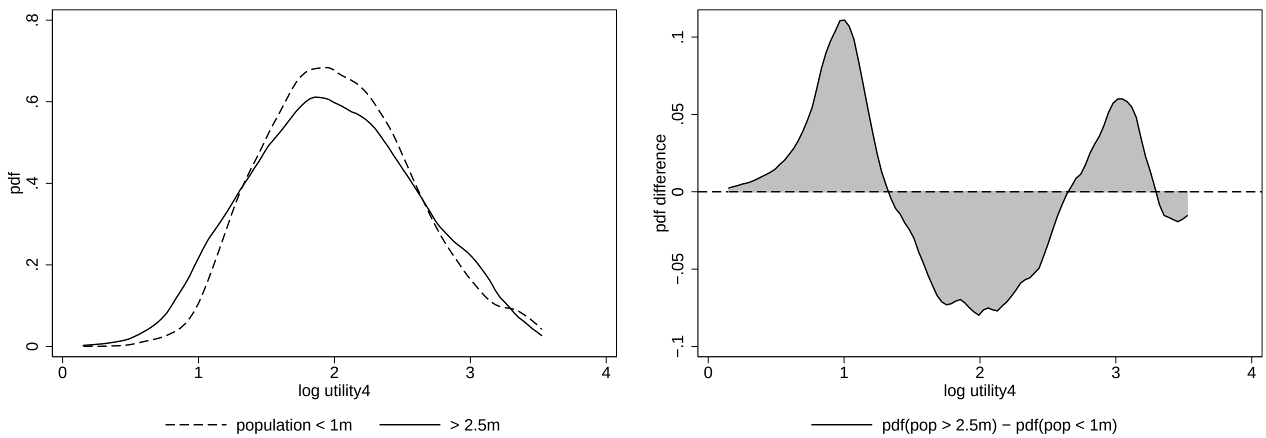

Figure 13: Skill distribution using ACCRA cost-of-living index, adjusting for variation in prices for all goods; Left graphs: density; Right graphs: Density differential.

the results of price differences across borders, Broda and Weinstein (2008) find that significant price differences that are found using price indexes are not replicated once they use US and Canadian barcode data. Their work is supportive of simple pricing models where the degree of market segmentation across the border is similar to that within borders. We therefore see panel A in Figure 13 as a very conservative upper bound of how the inclusion of consumption price differentials affects our initial findings.

7.5 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 14.

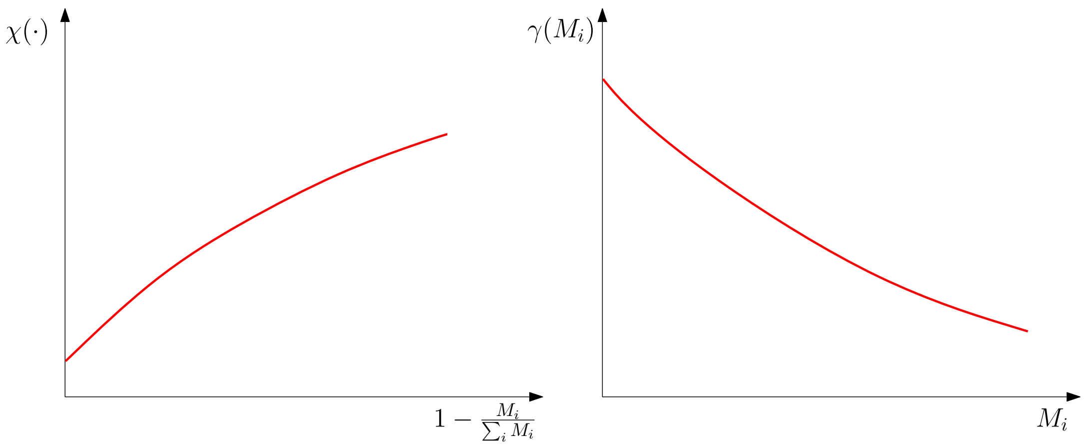

Figure 14: A. Spillover technology $\chi(\cdot)$, increasing in measure of other skills; B. The marginal product $\gamma(m_{ij})$ (and the Elasticity of Substitution ρ) are decreasing in abundance of skill.

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 σ_{ik} between skill i and k is given by (see Appendix for the derivation):

$$\sigma_{ik} = \frac{\gamma_i n_i^{\gamma_i} y_i + \gamma_k n_k^{\gamma_k} y_k}{\gamma_i n_i^{\gamma_i} y_i \left(1 - \gamma_k\right) + \gamma_k n_k^{\gamma_j} y_k \left(1 - \gamma_i\right)},$$

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 $\sigma = \frac{1}{1-\gamma}$. The technology with varying elasticity is compared to the constant elasticity technology in Figure 15.

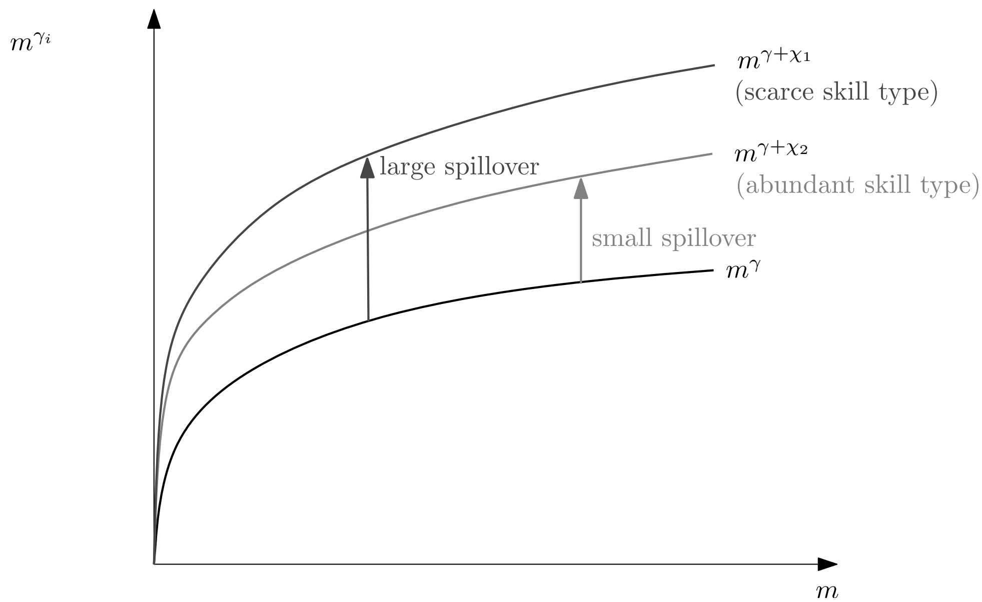

Figure 15: Scarce skill types are more likely to interact with agents with different skills and therefore, given m, they benefit from a larger expected spillover.

7.6 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.

Consideration of search frictions immediately brings up the issue of dynamics, currently completely absent in our analysis. Possibly cities of different size and TFP offer different earnings paths. Those with a steeper earnings path will induce workers to accept lower wages early on which over the entire path leaves them indifferent. This now potentially becomes a hairy dynamic problem. An important related empirical and modeling issue to be addressed in a dynamic framework is the age distribution across cities. Young people move into cities to move out again at middle age. This indicates that the benefits of large cities is non-monotonic over the life cycle which renders the dynamic implications of the model a priori ambiguous. Given the complexity of the issue, we leave dynamics for future work. Nonetheless, the approach adopted by Desmet and Rossi-Hansberg (2010) in solving for a dynamic spatial equilibrium will certainly be promising also here. They show that under certain assumptions on the distribution of property rights, such a complicated dynamic problem leads to static optimization. This also indicates that our static approach is appropriate.

Finally, yet another alternative explanation can be found in the division of labor. Based on a model of specialization, Duranton and Jayet (2010) find evidence using French data, that scarce occupations are overrepresented in larger cities.

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 – educational attainment – 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 fact that we find the same result when we use skill measures based on both observables and unobservables and measures based on observables only is indicative of the robustness of the result. 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 useful for many questions at hand, 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 equilibrium conditions (1) and (2) as:

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

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)–(4), with $H_1 = H_2$. Equating (3) and (4), we obtain:

$$\sum_{i=1}^{I} \left(\frac{M_i}{N_2}\right)^{\gamma_i} \frac{A_2}{p_2} \frac{(1-\gamma_i)\, y_i^{\beta}}{\left(1 + \frac{N_1}{N_2}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}}\right)^{\gamma_i}} \left\{\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_i}{\gamma_i-1}} - 1\right\} = 0 \qquad (6)$$

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:

$$\frac{A_1}{A_2}\left(\frac{p_2}{p_1}\right)\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{\gamma_i}{\gamma_i-1}} - 1 = \left(\frac{p_2}{p_1}\right)^{1-\alpha}\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}} - 1, \qquad (7)$$

and given $\frac{p_2}{p_1} < 1$, the equality in equation (6) 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.

Now, let's go back to equation (1)–(2) from the system:

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

Since we showed that $\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha} < 1$, we must have $\left[\frac{A_2}{A_1}\left(\frac{p_1}{p_2}\right)^{\alpha}\right]^{\frac{1}{\gamma_i-1}} > 1$, for every γ_i. This implies that LHS of the above expression is larger than 1. Assuming $H_1 = H_2$, this implies that:

$$\frac{p_1}{p_2}\frac{A_2}{A_1}\frac{N_2}{N_1} > 1$$

Rearranging, we have:

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

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

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

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:

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

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 $S_1 > S_2$. ■

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:

$$\begin{aligned} f_i &= A\gamma_i m_i^{\gamma_i-1} y_i^{\beta} \\ f_k &= A\gamma_k m_k^{\gamma_k-1} y_k^{\beta} \\ f_{ii} &= A\gamma_i \left(\gamma_i - 1\right) m_i^{\gamma_i-2} y_i^{\beta} \\ f_{kk} &= A\gamma_k \left(\gamma_k - 1\right) m_k^{\gamma_j-2} y_k^{\beta} \\ f_{ij} &= 0 \end{aligned}$$

Then, we have:

$$\begin{aligned} f_i m_i + f_k m_k &= A\gamma_i m_i^{\gamma_i} y_i^{\beta} + A\gamma_k m_k^{\gamma_k} y_k^{\beta} \\ f_i f_k &= A^2 \gamma_i \gamma_k m_i^{\gamma_i-1} m_k^{\gamma_k-1} y_i^{\beta} y_k^{\beta} \end{aligned}$$

Therefore, the numerator is:

$$A^3 \gamma_i \gamma_k y_i^{\beta} y_k^{\beta} m_i^{\gamma_i-1} m_k^{\gamma_k-1} \times \left(\gamma_i m_k^{\gamma_i} y_i^{\beta} + \gamma_k m_k^{\gamma_k} y_k^{\beta}\right)$$

and

$$\begin{aligned} f_i^2 f_{kk} &= A^3 \gamma_i^2 m_i^{2\gamma_i-2} y_i^{2\beta} \gamma_k \left(\gamma_k - 1\right) m_k^{\gamma_k-2} y_k^{\beta} \\ f_k^2 f_{ii} &= A^3 \gamma_k^2 m_k^{2\gamma_k-2} y_k^{2\beta} \gamma_i \left(\gamma_i - 1\right) m_i^{\gamma_i-2} y_i^{\beta} \end{aligned}$$

Then, the denominator is:

$$m_i m_k \times \left(\begin{array}{c} A^3 \gamma_i^2 m_i^{2\gamma_i-2} y_i^{2\beta} \gamma_k \left(\gamma_k - 1\right) m_k^{\gamma_k-2} y_k^{\beta} + \\ + A^3 \gamma_k^2 m_k^{2\gamma_k-2} y_k^{2\beta} \gamma_i \left(\gamma_i - 1\right) m_i^{\gamma_i-2} y_i^{\beta} \end{array} \right)$$

Rearranging, we have:

$$A^3 \gamma_i \gamma_k y_i^{\beta} y_k^{\beta} m_i^{\gamma_i-1} m_k^{\gamma_k-1} \times \left(\gamma_i m_i^{\gamma_i} y_i^{\beta} \left(\gamma_k - 1\right) + \gamma_k m_k^{\gamma_k} y_k^{\beta} \left(\gamma_i - 1\right)\right)$$

Therefore, the elasticity of substitution becomes:

$$\sigma_{ik} = -\frac{A^3 \gamma_i \gamma_k y_i y_k^{\beta} m_i^{\gamma_i-1} m_k^{\gamma_k-1} \times \left(\gamma_i m_i^{\gamma_i} y_i^{\beta} + \gamma_k m_k^{\gamma_k} y_k^{\beta}\right)}{A^3 \gamma_i \gamma_k y_i^{\beta} y_k^{\beta} m_i^{\gamma_i-1} m_k^{\gamma_k-1} \times \left(\gamma_i m_i^{\gamma_i} y_i^{\beta} \left(\gamma_k - 1\right) + \gamma_k m_k^{\gamma_k} y_k^{\beta} \left(\gamma_i - 1\right)\right)}$$

Simplifying, we get:

$$\sigma_{ik} = \frac{\gamma_i m_i^{\gamma_i} y_i^{\beta} + \gamma_k m_k^{\gamma_k} y_k^{\beta}}{\gamma_i m_i^{\gamma_i} y_i^{\beta} \left(1 - \gamma_k\right) + \gamma_k m_k^{\gamma_k} y_k^{\beta} \left(1 - \gamma_i\right)}$$

Notice that if $\gamma_i = \gamma_k = \gamma$, 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.[13] 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) [14]. 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.[15] 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). We also drop the lowest 0.5% of wages as a pragmatic way of eliminating likely misreported wages close to zero. 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.

Local house and commodity price indices

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).[16]

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.

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) [17] 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

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

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

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

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

[17] Available at http://mcdc2.missouri.edu/websas/geocorr2k.html.

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).[18]. 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.

For robustness checks, we also purchased the ACCRA Cost of Living Index from C2ER (The Council for Community and Economic Research). ACCRA data are collected by local chambers of commerce and similar organization who have volunteered to participate. They are reported for 269 core-based statistical areas (CBSA) and 80 metropolitan divisions for the 33 largest CBSAs. The ACCRA Cost of Living Index consists of six major categories: grocery items, housing, utilities, transportation, health care, and miscellaneous goods and services. These major categories in turn are composed of subcategories, each of which is represented by one or more items in the Index. In total, local prices of 60 items are reported, e.g. tbone steak (item 1), phone (31), gasoline (33), Lipitor (38), pizza (40) haircut (42), movie (52). Indices for major categories and an overall composite index are calculated as weighted averages where weights come from the Consumer Expenditures Survey conducted by the U.S. Bureau of Labor Statistics. We use the average of quarterly data from Q2.2008 to Q2.2009 in order to minimize the number of missing cities from non-reporting places. We use the average across metropolitan divisions to match ACCRA data to our wage data.

For further robustness checks, we use data from the National Association of Realtors for the 4th quarter in 2009. We use the median sales price of existing single-family homes for metropolitan areas. MSAs are as defined by the U.S. Office of Management and Budget and include the specified city or cities and surrounding suburban areas.

Firm size distribution

Data on the local distribution of firm sizes is taken from the county business patterns compiled by U.S. Census Bureau as an extract from its own Business Register (BR). Data for establishments are presented by geographic area, 6-digit NAICS industry, legal form of organization (U.S. only), and employment size class. Data consist of number of establishments, employment during the week of March 12, first quarter payroll, and annual payroll. We use aggregate data for 939 CBSAs in 2008.

[18] See http://www.fhfa.gov

Appendix C: Tables and Additional Figures

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
...		
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
254	Blacksburg-Christiansburg-Radford, VA	159,587
255	Bend, OR	158,629
256	Thomasville-Lexington, NC	158,582

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 256.

Table 2: Rank of cities by average log wages.

	City	Population	Avg. Log Wage
1	Bridgeport-Stamford-Norwalk, CT	901208	7.08
2	Barnstable Town, MA	221151	7.05
3	San Francisco-Oakland-Fremont, CA	4317853	6.99
4	Boulder, CO	303482	6.98
5	Ann Arbor, MI	347563	6.96
6	Washington-Arlington-Alexandria, DC-VA-MD-WV	5476241	6.95
7	Worcester, MA	803701	6.95
8	Oxnard-Thousand Oaks-Ventura, CA	802983	6.94
9	San Jose-Sunnyvale-Santa Clara, CA	1839700	6.93
10	Boston-Cambridge-Quincy, MA-NH	4588680	6.93
...			
247	Johnson City, TN	197381	6.38
248	Utica-Rome, NY	293280	6.38
249	Waco, TX	233378	6.37
250	Madera-Chowchilla, CA	148632	6.36
251	El Paso, TX	751296	6.35
252	Lynchburg, VA	247447	6.32
253	Jacksonville, NC	173064	6.31
254	Laredo, TX	241438	6.30
255	Amarillo, TX	246474	6.28
256	Brownsville-Harlingen, TX	396371	6.26

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				
1	–		–	
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				
1 year	–		–	
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	–		–	
1-family attached	-0.0805***	(0.0014)	-0.2059***	(0.0009)
2-family	-0.0875***	(0.0012)	0.0143***	(0.0015)
3-4 family	-0.1017***	(0.0012)	0.0158***	(0.0021)
5-9 family	-0.1068***	(0.0012)	-0.1578***	(0.0027)
10-19 family	-0.0550***	(0.0013)	-0.1982***	(0.0032)
20-49 family	-0.0865***	(0.0015)	-0.0803***	(0.0031)
50+ family	-0.0647***	(0.0013)	0.0452***	(0.0024)
not available	-0.0366***	(0.0028)	-0.0058***	(0.0010)
Bedroom to room ratio	0.1488***	(0.0027)	0.2824***	(0.0020)
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 groups are indicated by ”–”.

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	347563	4.88
2	Bridgeport-Stamford-Norwalk, CT	901208	4.88
3	Barnstable Town, MA	221151	4.87
4	Jackson, MI	159828	4.86
5	Worcester, MA	803701	4.85
6	Flint, MI	424043	4.84
7	Detroit-Warren-Livonia, MI	4403437	4.81
8	Boulder, CO	303482	4.81
9	Hartford-West Hartford-East Hartford, CT	1195998	4.79
10	Decatur, AL	151399	4.79
...			
244	Honolulu, HI	907574	4.35
245	Laredo, TX	241438	4.33
246	Brownsville-Harlingen, TX	396371	4.32
247	Utica-Rome, NY	293280	4.32
248	El Paso, TX	751296	4.31
249	Chico, CA	220577	4.30
250	Lynchburg, VA	247447	4.30
251	Madera-Chowchilla, CA	148632	4.28
252	Amarillo, TX	246474	4.25
253	Jacksonville, NC	173064	4.22

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

	City	Population	S.D. Utility
1	Santa Cruz-Watsonville, CA	256218	0.77
2	Punta Gorda, FL	156952	0.77
3	Boulder, CO	303482	0.71
4	San Jose-Sunnyvale-Santa Clara, CA	1839700	0.69
5	Springfield, OH	139671	0.69
6	Springfield, IL	208182	0.68
7	Bridgeport-Stamford-Norwalk, CT	901208	0.68
8	Lubbock, TX	276659	0.67
9	Durham-Chapel Hill, NC	501228	0.66
10	New York-Northern New Jersey-Long Island, NY-NJ-PA	19069796	0.66
...			
244	Myrtle Beach-North Myrtle Beach-Conway, SC	263868	0.43
245	Bellingham, WA	200434	0.42
246	Lynchburg, VA	247447	0.42
247	South Bend-Mishawaka, IN-MI	317538	0.42
248	Anderson, IN	131417	0.41
249	Macon, GA	231576	0.40
250	Appleton, WI	221894	0.40
251	Panama City-Lynn Haven-Panama City Beach, FL	164767	0.38
252	Janesville, WI	160155	0.37
253	Waco, TX	233378	0.37

Table 7: Regression coefficient on 2009 population size of average and standard deviation of real wage by skill category.

		average		standard deviation		N
Skill Category						
1	No high school	-0.011	(0.013)	0.042***	(0.008)	160
2	High school degree	-0.003	(0.009)	0.033***	(0.005)	249
3	Some college	0.003	(0.007)	0.022***	(0.004)	247
4	Bachelor	0.005	(0.009)	0.030***	(0.006)	237
5	Master	0.049***	(0.017)	0.039***	(0.011)	168
6	MD,...	-0.022	(0.044)	0.018	(0.034)	57
7	PhD	-0.023	(0.041)	0.015	(0.025)	55
All skill categories		0.016**	(0.007)	0.023***	(0.003)	254

Notes: 1. For many small cities we oberve less than two workers in the very low and very high education groups and cannot estimate the standard deviation. These city-group pairs are dropped; the samples on which the regressions are based do therefore systematically vary and the results are only indicative. 2. Standard error in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

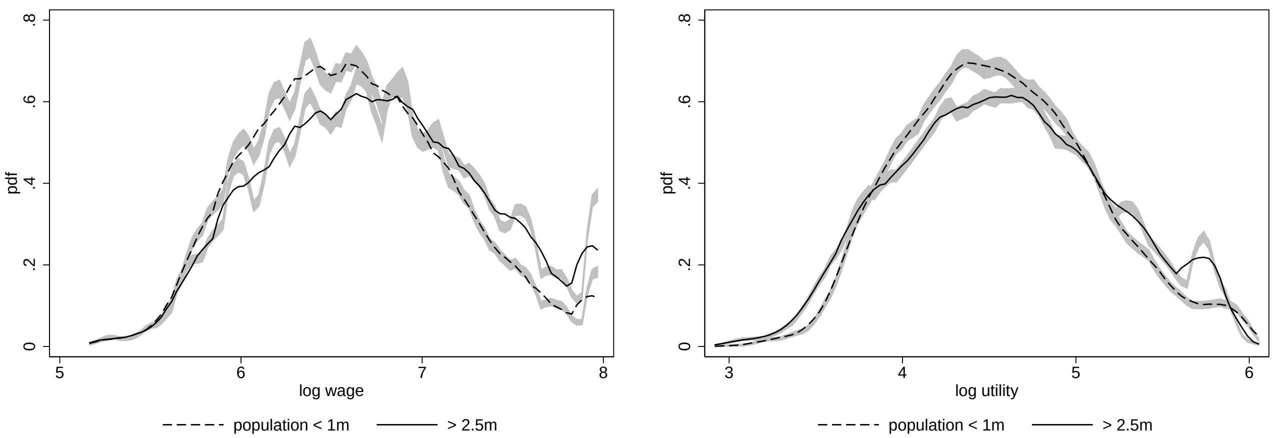

Figure 16: Wage and skill distribution for small and large cities with optimal band width and 95%-confidence intervals. Full-time wage earners from 2009 CPS. Kernel density estimtates (Epanechnikov kernel), not accounting for top-coding. Bandwidth point estimate: 0.06497 (small cities), 0.06690 (large cities); oversmoothed interval bandwidth: 0.03889 (small cities), 0.03965 (large cities).

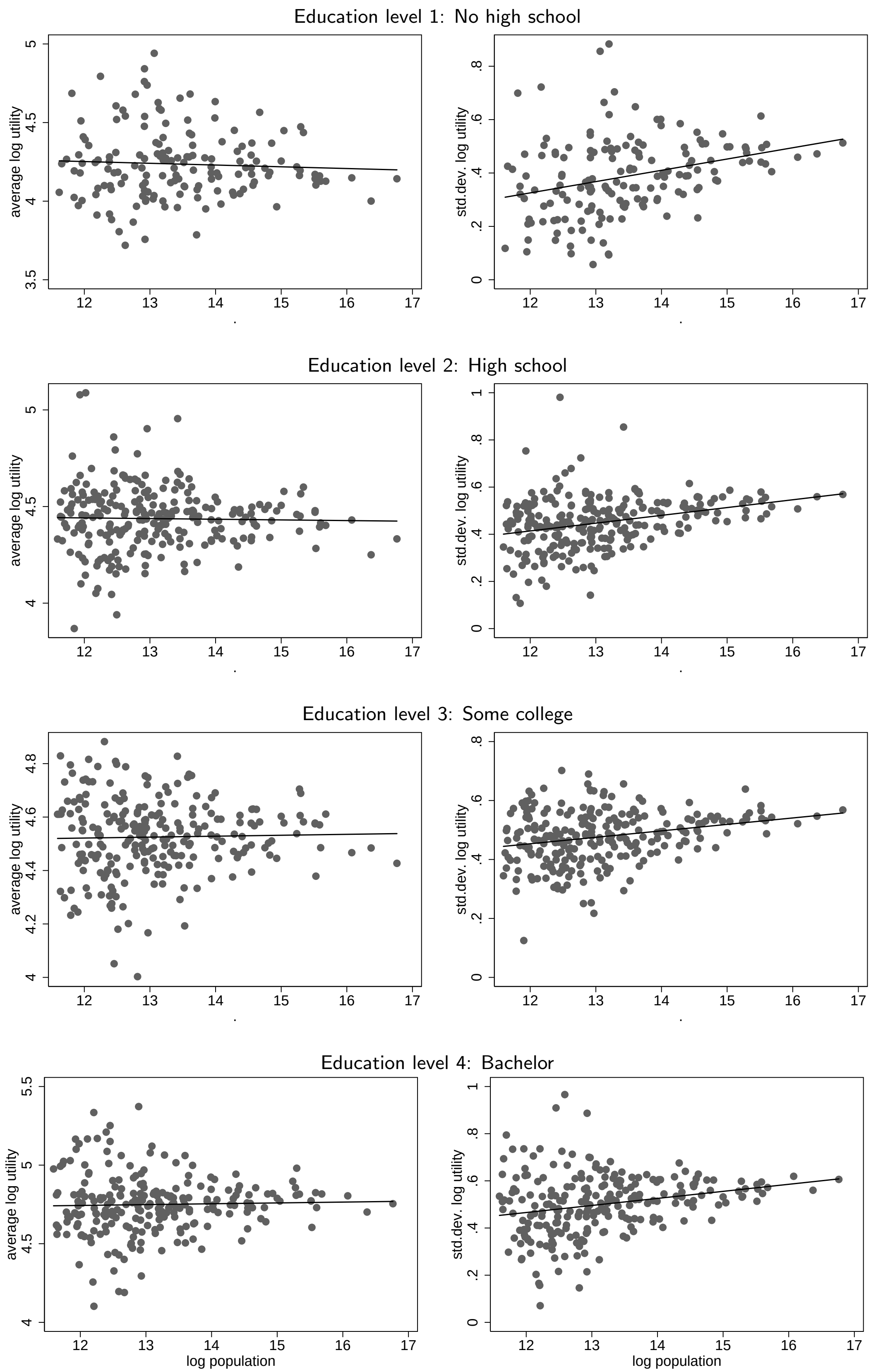

Figure 17: Skill distribution by population size conditional on 7 observed educational levels. Left graphs: Mean; Right graphs: Standard Deviation.

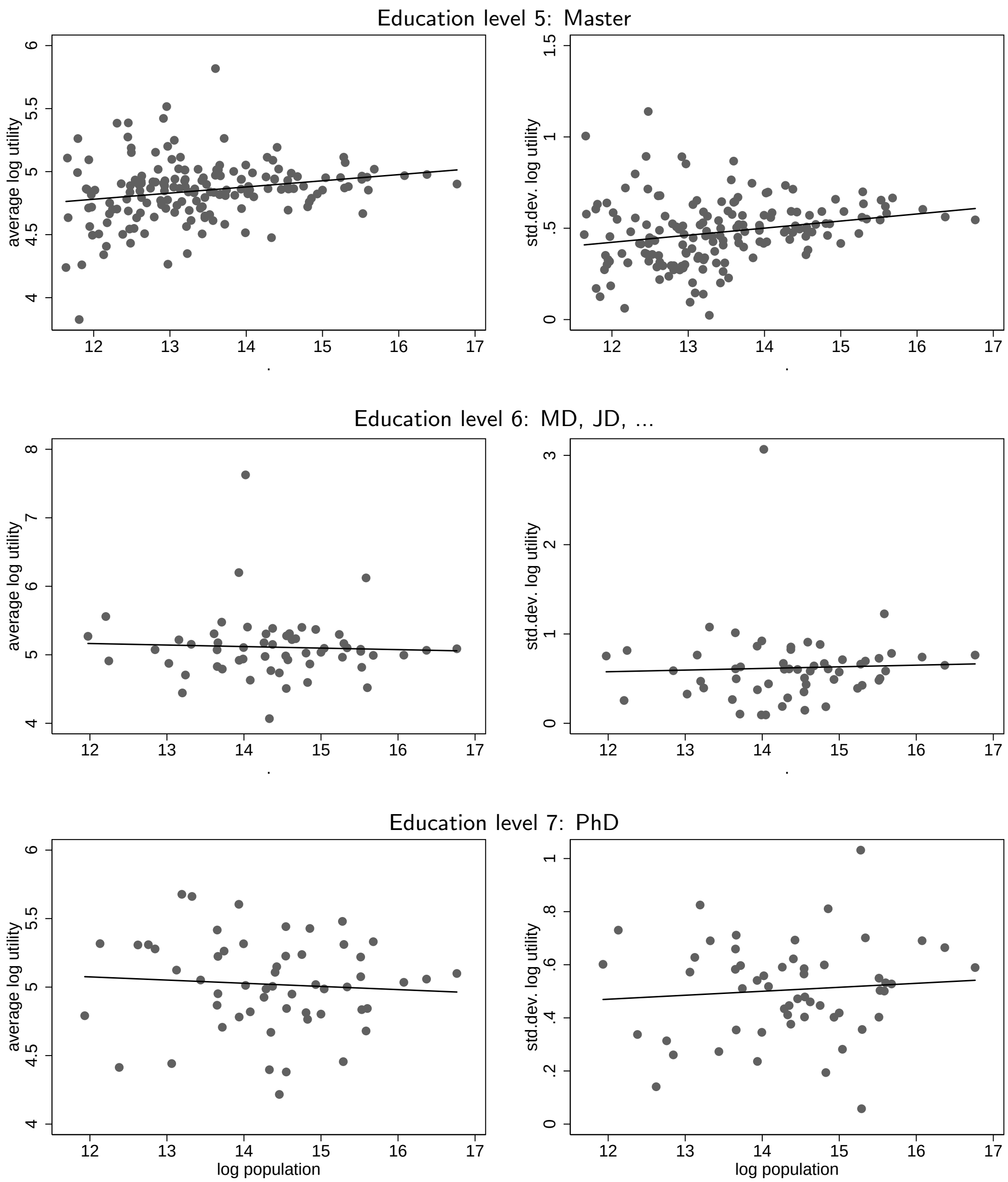

Figure 17: continued.

References

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

[2] Bacolod, Marigee, Bernardo Blum, and William Strange, "Skills in the City," *Journal of Urban Economics* 65(2), 2009, 136-153.

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

[4] Broda, Christian, and David E. Weinstein, "Understanding International Price Differences Using Barcode Data," NBER Working Paper #14017, May 2008.

[5] Broda, Christian, Ephraim Leibtag, and David E. Weinstein, "The Role of Prices in Measuring the Poor's Living Standards," *Journal of Economic Perspectives* **23(2)**, 2009, 7797.

[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] Desmet, Klaus, and Esteban Rossi-Hansberg, "Spatial Development," Princeton mimeo 2010.

[12] Duranton, Gilles, and Hubert Jayet, "Is the Division of Labour Limited by the Extent of the Market? Evidence from French Cities," forthcoming *Journal of Urban Economics* 2010.

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

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

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

[16] Fu, S., and Stephen Ross, "Wage Premia in Employment Clusters: Agglomeration or Worker Heterogeneity?" mimeo, 2010.

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

[18] 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.

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

[20] Helpman, Elhanan, Oleg Itskhoki, and Stephen Redding "Inequality and Unemployment in a Global Economy," *Econometrica*, **78 (4)**, 2010, 12391283.

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

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

[23] Koo, Jahyeoung, Keith R. Phillips, and Fiona D. Sigalla, "Measuring Regional Cost of Living.," *Journal of Business & Economic Statistics,* **18**, 2000, 127-136.

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

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

[26] 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.

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