Spatial Sorting*

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

February 24, 2014

Forthcoming in the *Journal of Political Economy*

Accepted December 27, 2013

Abstract

We investigate the role of skill complementarities in production and mobility across cities. The nature of the complementarities determines the equilibrium skill distribution across cities. With extreme-skill complementarity, the skill distribution has thicker tails in large cities; with top-skill complementarity, there is first-order stochastic dominance. Using wage and housing price data, we find robust evidence of thick tails in large cities: large cities disproportionally attract both high and low-skilled workers, while average skills are constant across city size. This pattern of spatial sorting is consistent with extreme-skill complementarity, where the productivity of high-skilled workers and of the providers of low skilled services is mutually enhanced.

Keywords. Complementarity. Cities. Sorting. Price-Theoretic Measure of Skills. Population Mobility. City Size. Matching Theory. General equilibrium. Skill Distribution.

*We are grateful to numerous colleagues for detailed discussion and insightful comments. We also benefited from the feedback of many seminar audiences. Eeckhout gratefully acknowledges support by the ERC, Grant 208068 and Schmidheiny by the Swiss National Science Foundation, Grant Sinergia/130648.

[‡]University College London and GSE-UPF, jan.eeckhout@ucl.ac.uk.

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

[¶]University of Basel and CEPR, kurt.schmidheiny@unibas.ch.

"*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

Complementarities are important for the productivity and composition of workers in firms, for student achievement in classrooms (peer effects) and for the accomplishments of teams. The presence of more productive co-workers affects the performance of some if not all other co-workers, and this in turn determines who chooses to work where and with whom. In this paper we investigate the role of complementarities at an aggregate level – the level of a city. Complementarities are akin to knowledge spillovers (Marshall, 1890), but rather than flowing between innovating firms, complementarities affect the productivity of *differentially* skilled workers within the local labor market. We propose a model that elucidates both the nature of cities and the role of complementarities in production. Our main theoretical finding is that the specifics of the complementarities determine the distribution of skills within a city and how it varies with city size. Our approach sheds new light on the sources of the urban wage premium, a major puzzle in the literature. It is well known that wages in large cities are higher, but it is unclear why. Little is known about the skill composition across cities. Are wages higher because workers in large cities are more skilled? Most people can provide casual evidence that the skill level in the top percentiles of New York and large cities in general is higher than anywhere else. Making it there – in New York, NY – rather than in Akron, OH is the ultimate aim of many professionals in many trades: artists, musicians, advertising and media professional, consultants, lawyers, financiers, ... Yet, cities are not just populated by superstars and high earning professionals, even if these are highly visible.

In this paper, we address the sorting decision of workers over the entire range of skills, including medium and low skills. Our main empirical and hitherto undocumented finding is that the distribution of skills in the US has thick tails in large cities: large cities disproportionally attract both high and low-skilled workers, while average skills are constant across size. From the theory, this allows us to conclude that there are complementarities between high and low-skilled workers, which mutually boosts their productivity.

We consider two competing hypotheses concerning the complementarities between skills. A first hypothesis is that the superstars boost their productivity most in the presence of other high-skilled workers. For example, under this assumption the best lawyers are more productive when surrounded by top legal assistants. Or the cancer surgeons at Sloan Kettering in NY work best with top residents and top nurses, whereas the General Practitioners with fewer years of training and fewer fellowships collaborate with less trained nurses and assistants. We refer to this hypothesis as *Top-Skill Complementarity*. A second hypothesis posits that high-skilled workers boost their productivity most with low-skilled services. What dominates in the aggregate is that the high-skilled worker has a disproportionately high productivity increase from the presence of low-skilled services. Given the value of her time, at her job she hires more low-skilled administrative help, other services (sales, legal, catering...) and she demands low-skilled services through child care, schooling and help in the household. We label this hypothesis as *Extreme-Skill Complementarity*.

The premise of our analysis is that the presence of those complementarities determines the location decision of differentially skilled workers, i.e. spatial sorting. We propose a theory that identifies a one to one relationship between those features of the technology, on the one hand, and the equilibrium outcome of the skill distributions across cities, on the other. Our model is a tractable version of the multi-worker matching model à la Kelso and Crawford (1982) applied to a concrete labor market setting. Complementarities determine competitive wages and therefore the location decision of workers. Our objective is to uncover the nature of the complementarities from the observed sorting pattern of workers, i.e. the skill distribution across cities. This is very much in the spirit of Krusell, Ohanian, Rios-Rull and Violante (2000) in the macro literature, who derive properties of complementarities in technology from the observed wage distributions.

Our labor market model is tailored to investigate the nature of cities, and the *contribution* of the paper is threefold: First, we identify a mechanism of skill complementarities and the resulting skill distribution that we can explicitly solve. This, despite the fact that models with varying elasticity of substitution are notoriously hard to solve analytically, as Krusell, Ohanian, Rios-Rull and Violante's (2000) dynamic model illustrates. Second, qualitatively we discover an extremely robust empirical pattern of thick tails in the distribution of skills: average skills are independent of city size, while the standard deviation increases with city size. In conjunction with the results from the theory, this allows us to conclude that the observed pattern of skills is due to the complementarities between extreme skills. We believe that the theoretical link between the complementarities and the distribution in local labor markets is both theoretically and empirically novel. Third, our analysis makes further headway in our understanding of one of the major outstanding puzzles in urban economics, namely the mechanism behind the urban wage premium. Our findings establish that wages are not higher because skills are uniformly higher. In fact, average skills are constant across cities. Wages are therefore higher only to compensate for higher housing prices. But our findings also show that the skill composition is important for the distribution of wages and productivity. Complementarities between extreme skills act as a multiplier of existing differences in total factor productivity across cities. Our theory of differential complementarities provides an explanation for higher wage and skill inequality in large cities.

The normative implications in the baseline model are particularly relevant when evaluating inequality. Our results show that from a social welfare viewpoint, wage inequality and urbanization are intimately related. This can have far-reaching policy implications. Consider for example the current income taxation system that progressively taxes individuals and households on their nominal income. Effectively, given the urban wage premium, this means that the current fiscal system differentially taxes cities of different sizes. The implication is that the population distribution across different size cities is distorted, and as a result, aggregated output produced is suboptimal.

Prices play a key role in our equilibrium model of city choice. Heterogeneously-skilled citizens earn a living based on a competitive wage and choose housing in a competitive housing market. Under perfect mobility, their location choice will make them indifferent between consumption-housing bundles, and therefore between different wage-housing 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). This naturally gives rise to a *price-theoretic measure of skills*. Larger cities pay higher wages, and are more expensive to live in. Under worker mobility, revealed preference location choices imply that wages adjusted for housing prices are a measure of skills.

Using this price based measure of skills, we can establish two robust empirical facts: average skills are constant across cities, and the standard deviation increases with city size. Big cities are characterized by big real inequality. The city size-wage premium is thus not driven by a high average skill level. Instead, larger cities have thicker tails in the skill distribution and disproportionately attract both higher and lower skilled agents. In New York City for example there is not only a huge contingent of high-skilled workers in Manhattan, but there are also disproportionately many low-skilled workers living in the South Bronx and Newark. 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 they are to small cities because of the systematic pattern of thick tails in the skill distribution of large cities.

We document that this systematic pattern of spatial sorting is extremely robust to different measures: we use educational attainment and occupation as direct measures of skills and control for industry selection, we investigate the role of migration, we consider different definitions of large versus small

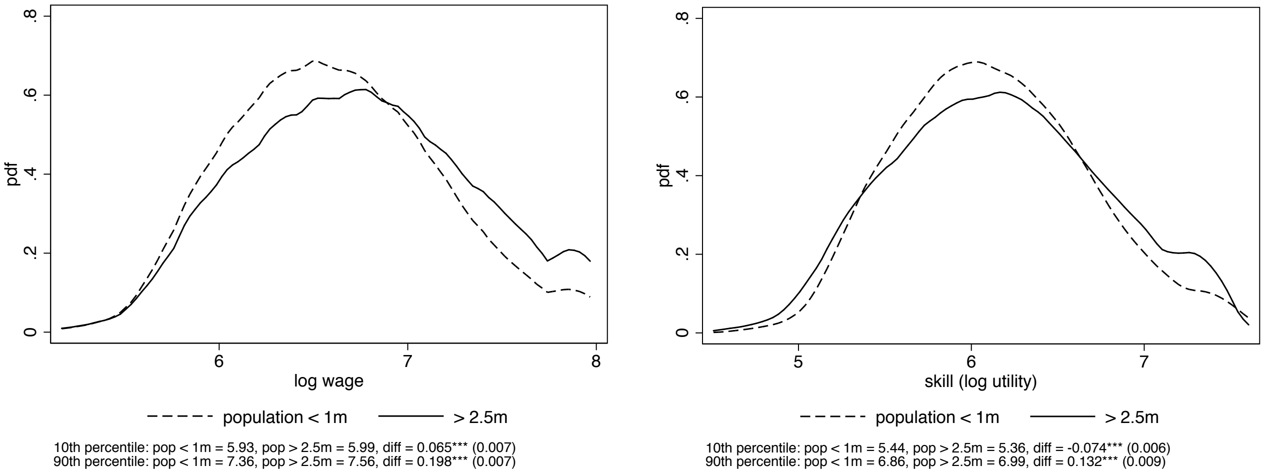

Figure 1: Wage and skill distribution for small and large cities, Kernel density estimates (Epanechnikov kernel, bandwidth = 0.1); Wage data from the 2009 Current Population Survey (CPS) on 25,584 workers in 202 small CBSAs (population between 100,000 and 1m) and 34,999 workers in 21 large CBSAs (larger 2.5m); A. Wages; B. Skills. Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

cities, we use three different data sources for local housing values, and we include local price differences in consumption goods. To our knowledge, this pattern of spatial sorting – that mobility across cities driven by differential skill complementarity determines the skill distribution – has not been documented in the literature.

The thick tails in the skill distribution are nonetheless consistent with the well-documented city-size wage premium. The gap between average wages in the smallest cities in our sample (with a population of around 160,000, more than 100 times smaller than New York) and the largest cities is 25%. 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 in, so in order to be able to compare skill distributions, we need to adjust for housing 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 housing price index based on a hedonic regression to calculate the difference in housing values across cities. Figure 1.B displays the kernel of the induced 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 finding sheds light on the nature of the underlying technology: the fat tails result is consistent with Extreme-Skill Complementarity.

A key feature of our approach is the price-theoretic measure of skills which allows us to characterize a smooth distribution of skills. This is in contrast to the common approach of using observable skills such as attained education levels or years of education. To investigate the role of observables in the spatial sorting pattern, we decompose the difference in the skill distribution between large and small cities. We find an asymmetry: in the lower tail, virtually none of the city size difference is explained by observables, while in the upper tail about half is explained, mainly by education. The high-skilled are more educated in large cities than in small cities, while the low-skilled are equally educated across city sizes.

There are of course other possible alternative explanations. And while we cannot exhaustively analyze all alternatives empirically, we can rule out a few prominent candidate alternatives and establish the robustness of our findings with respect to industry composition, migration and age or life cycle patterns. We discuss competing theoretical explanations such as the role of home versus market production in a world where agents have preferences for low-skilled services. We also investigate the role of non-homothetic preferences and within-city sorting.

2 Related Literature

The model we propose builds on the urban location model in Eeckhout (2004) and Davis and Ortalo-Magné (2009) (see also Guerrieri, Hartley, and Hurst (2011), who augment the model with local externalities) where identical citizens who have preferences over consumption and housing choose a city in order to maximize utility. This model has been used to explain population dynamics (see also Gabaix, 1999) and expenditure shares. Because of differences in productivity across cities, wages differ and housing 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. Our main innovation over the existing model is the introduction of heterogeneity in the inputs of production (skills) which gives rise to a distribution of skills within the city. This is necessary to meaningfully address sorting of heterogeneous agents within and across cities. Technology allows for varying degrees of complementarities between different skill types. Equilibrium is determined by the sorting decision of agents.

In recent work, Behrens, Duranton and Robert-Nicoud (forthcoming) analyze the distribution of heterogeneous agents across cities. Their model predicts perfect sorting by talent: New York attracts all the Ph.D.'s, Los Angeles and Chicago all the Masters, ..., and all the high school dropouts locate in small cities like Janesville, WI. Within-city heterogeneity in productivity is due to an ex post shock upon which workers cannot relocate any more. As a result, they postulate first order stochastic dominance of the (degenerate) talent distributions rather than the thick tails that our model predicts and that we find in the data. Their result of perfect sorting in talent is a direct consequence of assuming that each worker consumes one unit of land independent of his wage. Highly talented workers with high wages are therefore, relatively, much less affected by high housing costs in large cities than less talented workers. In contrast, we do allow housing consumption to increase with wages and hence with talent in line with overwhelming empirical evidence. In the equilibrium allocation of our model, *all* skill types locate in *all* cities simultaneously, driven by complementarities in production. This not only gives rise to a wage and skill distribution with full support as observed across cities of all sizes. Most importantly, it also allows us to infer the pattern of complementarities in the technology that drive the location decision of workers.

Finally, Van Nieuwerburgh and Weill (2010) set up a spatial equilibrium model to explain the increase in housing price dispersion resulting from an increase in productivity dispersion of heterogeneous workers. As in our model, worker mobility in response to productivity shocks and endogenous housing prices are the main ingredients of their explanation. The key difference is the production function for consumption goods which is linear in labor and does not feature skill complementarities.

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. 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 when local differences in federal taxes, non-labor income and observable amenities such as seasons, sunshine, and coastal location are controlled for, quality of life does not depend on size.

This body of work is consistent with our finding that the average of the skill distribution is remarkably constant across different sized cities. Of course, that does not allow us to conclude that there is no sorting of high-skilled workers into large cities and of low skilled workers into small cities. As we will show below, the data reveals quite the contrary. The mean is constant across cities of different size, but the variance increases substantially. The latter indicates an important role for sorting of high and low skill types into large cities and of medium types into small cities.

Our findings are also related to the previous literature on variations in the measured 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. However they are not 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 in large cities for the lowest scores (10th percentile) are much lower than the ones in small cities. Similarly, the highest scores (90th percentile) are much higher in large cities than in small ones, which is consistent with our robustness exercise reported in section 7.1 on direct measures of education. Also Gautier and Teulings (2009, Table 1) report a higher measured standard deviation in educational attainment across larger cities, which is consistent with our findings. However, they find first order stochastic dominance rather than thick tails as we do, because the mean is also higher. The reason is that their measure of skill is really a measure of wages. It is independent of housing prices and it is constructed as predicted wages net of unobserved heterogeneity using a Mincerian wage regression. Consistent with the urban wage premium, average wages increase with city size. Instead, our measure of skills adjusts wages for the equilibrium mobility decision by means of housing prices and we find that average skills are independent of city size. Together with the fact that the standard deviation increases with city size, this gives us thick tails, and not first order stochastic dominance. Note also that the first order stochastic dominance in Gautier and Teulings (2009) is not consistent with the direct measures of skills as reported in section 7.1 or as documented by Bacolod, Blum and Strange (2009). The distributions of those direct measures have thick tails in large cities – equal means, higher standard deviation – just as our wage-based measure that is adjusted for housing prices. In sum, while the *wage* distribution has been shown to satisfy first order stochastic dominance in city size (increasing mean, increasing standard deviation), we establish that the distribution of our wage-based measure of *skills* has thick tails in large cities (equal mean, increasing standard deviation).

There is also recent literature on increasing wage inequality over time.[1] Autor and Dorn (2013)

[1] Moretti (2013), Baum-Snow and Pavan (2012) and Autor and Dorn (2013).

document faster growth at both tails of the wage distribution between 1980 and 2005 and attribute this to the falling cost of automating middle skill routines, the polarization hypothesis.

Finally, there is little direct evidence on the role of complementarities between heterogeneously-skilled agents. One notable exception is the work by Hamilton, Nickerson and Owan (2003), who analyze the effect of team composition on productivity in a textiles production plant. They find that heterogeneous teams are more productive with average productivity held constant. While their setup is very specific and other theories can certainly rationalize this outcome, their finding is consistent with a technology that has extreme-skill complementarity.

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: $i \in \mathcal{I} = \{1, ..., I\}$. Associated with this skill order is a level of productivity y_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 tradable 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.[2] A worker has consumer preferences over the quantities of goods and housing c and h that are represented by: $u(c, h) = c^{1-\alpha}h^{\alpha}$, where $\alpha \in [0, 1]$. Workers are perfectly mobile, so they can relocate instantaneously and at no cost to another city. Because workers with the same skill are identical, in equilibrium each of them should obtain the same utility level wherever they choose to locate. Therefore for any two cities j, j' it must be the case that the respective consumption bundles satisfy $u(c_{ij}, h_{ij}) = u(c_{ij'}, h_{ij'})$, for all skill types $\forall i \in \{1, ..., I\}$.

Technology. Cities differ in their total factor productivity (TFP) which is denoted by A_j. For now, we assume that TFP is exogenous. We think of it as representing a city's productive amenities, infrastructure, historical industries, persistence of investments, etc.[3]

[2] 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 absentee landlords.

[3] In an earlier version of the paper, we endogenize A_j and let it be the result of agglomeration externalities. This is also documented in the Additional Material Section.

In each city, there is a technology operated by a representative firm that has access to a city-specific TFP A_j. Output is produced by choosing the right mix of differently skilled workers i. For each skill i, a firm in city j chooses a level of employment m_{ij} and produces output: $A_j F(m_{1j}, ..., m_{Ij})$. Firms pay wages w_{ij} for workers of type i. It is important to note that wages depend on the city j because citizens freely locate between cities not based on the highest wage, but, given housing price differences, based on the highest utility. Like land, firms are owned by absentee capitalists (or equivalently, all citizens own an equal share in the mutual fund that owns all the land and all the firms).

Market Clearing. In the country-wide market for skilled labor, markets for skills clear market by market, and for housing, there is market clearing within each city:

$$\sum_{j=1}^{J} C_j m_{ij} = M_i, \ \forall i \qquad\qquad \sum_{i=1}^{I} h_{ij} m_{ij} = H_j, \ \forall j. \tag{1}$$

where C_j denotes the number of cities with TFP A_j.

4 The Equilibrium Allocation

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

$$\begin{aligned} \max_{\{c_{ij},h_{ij}\}} u(c_{ij}, h_{ij}) &= c_{ij}^{1-\alpha} h_{ij}^{\alpha} \\ \text{s.t. } c_{ij} + p_j h_{ij} &\leq w_{ij} \end{aligned} \tag{2}$$

for all i, j. Solving for the competitive equilibrium allocation for this problem we obtain $c_{ij}^{\star} = (1-\alpha) w_{ij}$ and $h_{ij}^{\star} = \alpha \frac{w_{ij}}{p_j}$. Substituting the equilibrium values in the utility function, we can write the indirect utility for a type i as:

$$U_i = \alpha^{\alpha} (1-\alpha)^{1-\alpha} \frac{w_{ij}}{p_j^{\alpha}} \quad \Longrightarrow \quad w_{ij} = U_i p_j^{\alpha} \frac{1}{\alpha^{\alpha} (1-\alpha)^{1-\alpha}}, \tag{3}$$

where U_i is constant across cities from labor mobility. This allows us to link the wage distribution across different cities j, j'. Wages across cities relate as:

$$\frac{w_{ij}}{w_{ij'}} = \left(\frac{p_j}{p_{j'}} \right)^{\alpha}. \tag{4}$$

The Firm's Problem. All firms are price-takers and do not affect wages. Wages are determined simultaneously in each submarket i, j. Given the city production technology, a firm's problem is given by:

$$\max_{m_{ij},\forall i} A_j F(m_{1j}, ..., m_{Ij}) - \sum_{i=1}^{I} w_{ij} m_{ij}, \tag{5}$$

subject to the constraint that $m_{ij} \geq 0$. The first-order condition is: $A_j F_{m_{ij}}(m_{ij}) = w_{ij}, \forall i$.[4]

Because there is no general solution for the equilibrium allocation in the presence of an unrestricted technology, we focus on variations of the Constant Elasticity of Substitution (CES) technology, where the elasticity is allowed to vary across skill types. As a benchmark therefore, we consider the CES technology:

$$A_j F(m_{1j}, ..., m_{Ij}) = A_j \left(\sum_{i=1}^{I} m_{ij}^{\gamma} y_i \right) \tag{6}$$

with $\gamma < 1$. In this case the first-order conditions are $A_j \gamma m_{ij}^{\gamma-1} y_i = w_{ij}, \forall i$.

Below we will solve the allocation under CES as a special case of the more general technology. 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 is governed by two components: (1) the productivity y_i of the skilled labor and how fast it increases in i; and (2) the measure of skills m_{ij} employed (wages decrease in the measure employed from the concavity of the technology). Without loss of generality, we assume that wages are monotonic in the order i.[5] This is consistent with our price-theoretic measure of skill.

We now proceed by introducing varying degrees of complementarities between different skills, starting from the CES technology. This implies the technology now has an elasticity of substitution that is no longer constant. For tractability, let there be two cities, $j \in \{1, 2\}$ and three skill levels $i \in \{1, 2, 3\}$. Consider any subset of the skills, say i, k, between which there is a degree of complementarity λ, and none with the remaining skill level l. Then the technology can be written as $\left(m_{ij}^{\gamma} y_i + m_{kj}^{\gamma} y_k\right)^{\lambda} + m_{lj}^{\gamma} y_l$. Depending on the subset of skills, we distinguish between the following configurations.

[4] In what follows, the non-negativity constraint on m_{ij} is dropped. This is justified whenever the technology satisfies the Inada condition that marginal product at zero tends to infinity whenever A_j is positive. This will be the case since we focus on variations of the CES technology.

[5] For a given order i, wages may not be monotonic as they depend on the relative supply of skills as well as on y_i. If they are not, we can relabel skills such that the order i corresponds to the order of wages. Alternatively, we can allow for the possibility that higher skilled workers can perform lower skilled jobs. Workers will drop job type until wages are non-decreasing. Then the distribution of workers is endogenous, and given this endogenous distribution, all our results go through. For clarity of the exposition, we will assume that the distribution of skills ensures that wages are monotonic.

Definition 1 *Consider the following technologies:*

I. ***Extreme-Skill Complementarity.*** *High skill workers are complementary with low skill workers.*

$$A_j F(m_1, m_2, m_3) = A_j \left[\left(m_{1j}^{\gamma} y_1 + m_{3j}^{\gamma} y_3 \right)^{\lambda} + m_{2j}^{\gamma} y_2 \right], \tag{7}$$

when $\lambda > 1$ relative to CES. Instead, skills 1 and 3 are substitutes when $\lambda < 1$.

II. ***Top-Skill Complementarity.*** *High skill workers are complementary with medium skill workers.*

$$A_j F(m_1, m_2, m_3) = A_j \left[\left(m_{2j}^{\gamma} y_2 + m_{3j}^{\gamma} y_3 \right)^{\lambda} + m_{1j}^{\gamma} y_1 \right], \tag{8}$$

when $\lambda > 1$ relative to CES. Instead, skills 2 and 3 are substitutes when $\lambda < 1$.

Observe that we could also introduce bottom-skill complementarities. In terms of the distributional implications, this is equivalent to top-skill substitutabilities, i.e., technology II with $\lambda < 1$. There are therefore 5 distinct configurations of the technology: two for technology I., with complements ($\lambda > 1$) or substitutes ($\lambda < 1$), two for technology II. ($\lambda > 1$ and $\lambda < 1$), and CES ($\lambda = 1$).

It is worth pointing out that for our purpose, three skills is the minimal requirement to fully capture first order stochastic dominance and thick tails. Distinguishing between the two cannot be achieved with two skills only. At the same time, with a larger number of skills, we do not obtain qualitatively different results. With one hundred skill types, one can of course analyze the properties of each percentile, but that does not provide essential additional information about the existence of thick tails or stochastic dominance. We nonetheless investigate the generality of this setup. In the Online appendix we report the same properties that we derive below for general technologies with any N skills, and for more general patterns of gross complementarities. While we can handle a large number of cities, for analytical purposes, we cannot generalize the setup beyond two city types. That is, we can compute the equilibrium allocation[6], but we cannot find an analytical solution for it. We can however analyze the setting for *any* number of cities with types A_1 or A_2, i.e., for any C_1, C_2.

We first derive the equilibrium conditions for case I, Extreme-Skill Complementarity. The first-order

[6]Here it is worth drawing a parallel to the work by Krusel, Ohanian, Rios-Rull and Violante (2000). They compute an infinite (or long finite horizon) economy with intertemporal prices. The parallel to their dynamic economy is our cross section of cities with spatial prices.

conditions are for each j and all skill types i, respectively:

$$\lambda A_j \left[m_{1j}^{\gamma} y_1 + m_{3j}^{\gamma} y_3\right]^{\lambda-1} \gamma m_{1j}^{\gamma-1} y_1 - w_{1j} = 0 \tag{9}$$

$$\gamma A_j m_{2j}^{\gamma-1} y_2 - w_{2j} = 0 \tag{10}$$

$$\lambda A_j \left[m_{1j}^{\gamma} y_1 + m_{3j}^{\gamma} y_3\right]^{\lambda-1} \gamma m_{3j}^{\gamma-1} y_3 - w_{3j} = 0 \tag{11}$$

Using labor mobility, we can write the wage ratio in terms of the house price ratio for all i, $\frac{w_{i2}}{w_{i1}} = \left(\frac{p_2}{p_1}\right)^{\alpha}$ and equate the first-order condition in both cities for a given skill, for example for $i = 1$:

$$A_1 \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda-1} m_{11}^{\gamma-1} = \left(\frac{p_1}{p_2}\right)^{\alpha} A_2 \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda-1} m_{12}^{\gamma-1} \tag{12}$$

Using market clearing, $m_{12} = \frac{M_1}{C_2} - \frac{C_1}{C_2} m_{11}$ in the local labor market, we can solve for the first-order conditions for each skill to obtain the equilibrium quantities:

$$m_{11} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}} \frac{M_1}{C_2}}{1 + \frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}}, \quad m_{21} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma-1}} \frac{M_2}{C_2}}{1 + \frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\gamma-1}}}, \quad m_{31} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}} \frac{M_3}{C_2}}{1 + \frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}}, \tag{13}$$

and likewise in city 2.

So far we have consumer optimization for consumption and housing, the location choice by the worker, and firm optimization given wages. The next step is to allow for market clearing in the housing market given land prices. The system is static and solved simultaneously, which is reported in the Appendix. In what follows, we assume $H_j = H$ for all cities j. Below, we will discuss the implications where this simplifying assumption has bite.

The Main Theoretical Results. First we establish the relation between TFP and city size. Denote by S_j the size of city j where $S_j = \sum_{i=1}^{I} C_j m_{ij}$. When cities have the same amount of land, we can establish the following result.

Theorem 1 *City Size and TFP. Let $A_1 > A_2$ and $\lambda\gamma < 1, \gamma < 1$. Then the more productive city is larger, $S_1 > S_2$.*

Proof. In Appendix. ■

We establish this result for cities with an identical supply of land. Clearly, the supply of land is important in our model since in a city with an extremely small geographical area, labor demand would drive up housing prices all else equal. This may therefore make it more expensive to live in 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), we believe that this assumption does not lead to much loss of generality.[7]

We now establish the main theorem characterizing the skill distribution across firms. We already know that more productive cities are larger, but this 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.

Theorem 2 *Extreme-Skill Complementarity and thick tails. Given* $A_1 > A_2$*,* $\lambda > 1$ *and* $\lambda\gamma < 1$*, the skill distribution in the larger city has thicker tails.*

Proof. In Appendix. ■

Two corollaries immediately follow from the main theorem.

Corollary 1 *CES technology. If* $\lambda = 1$ *and* $\gamma < 1$*, then the skill distribution across cities is identical.*

Under CES technology, cities have identical skill compositions. This is due to the homotheticity of 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. The proof of the result follows immediately from setting $\lambda = 1$ in the proof of Theorem 2. Even though the city skill distribution under CES technology is the same across cities, the more productive city will be larger. This follows from Theorem 1.

The second Corollary establishes the mirror-image result under extreme-skill substitutability.

Corollary 2 *Extreme-Skill Substitutability and Thin Tails. Given* $A_1 > A_2$*,* $\lambda < 1$ *and* $\lambda\gamma < 1$*, the skill distribution in the larger city has thinner tails.*

These two corollaries can help build intuition for the result in Theorem 2. Consider first CES as a benchmark. 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 housing prices affect all skills within a city in the same way, the wage ratio is unaffected.

[7] In fact, the equal supply of housing condition is only sufficient for the proof, but not necessary. However, our model does not address 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. In Section 6.1, we empirically analyze the implications of within city sorting and find no qualitative impact on the results. This is consistent with recent work by Fu and Ross (2010), who find little evidence of sorting within metropolitan areas based on agglomeration.

Instead with extreme-skill complementarity, the marginal product of both the low and the high-skilled workers is higher than for medium skills, thus breaking the homotheticity. Given the complementarity between TFP A_j and the skill aggregator, the marginal impact on productivity of the extreme skills will now be disproportionately higher in larger than in smaller cities. This induces the relative increase in demand for extreme skills. Observe that this cannot be offset by higher housing prices because these 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 are equalized. This in-migration of low and high-skilled workers leads to the thick tails in the large cities.

Top-Skill Complementarity. Now consider the technology $A_j F(m_1, m_2, m_3) = A_j \left[\left(m_{2j}^{\gamma} y_2 + m_{3j}^{\gamma} y_3 \right)^{\lambda} + m_{1j}^{\gamma} y_1 \right]$. Without going through the detailed analysis in the text, we obtain the equivalent to Theorem 2 above (Theorem 1 readily extends as well) :

Theorem 3 *Top-Skill Complementarity and First Order Stochastic Dominance. Given $A_1 > A_2$, $\lambda > 1$ and $\lambda\gamma < 1$, the skill distribution in the larger city first-order stochastically dominates.*

Proof. In Appendix. ∎

And the corollary establishing the mirror-image result under extreme-skill substitutability.

Corollary 3 *Top-Skill Substitutability and First Order Stochastic Dominance. Given $A_1 > A_2$, $\lambda < 1$ and $\lambda\gamma < 1$, the skill distribution in the larger city is first-order stochastically dominated.*

Under top-skill complementarity, the highest skilled are complements with the next highest skill types, thus generating disproportionately higher output in larger cities. This complementarity breaks the homotheticity property, and leads to disproportionate demand in larger cities. Free mobility and real wage equalization across cities implies that the distribution in the larger city has disproportionately more of the top skill types. This induces first order stochastic dominance.

In Theorems 2 and 3 we identify a mechanism of skill complementarities in the production technology that generates a systematic pattern in the skill distribution. There is exactly one distribution pattern that corresponds to each of the 5 technology patterns (extreme-skill and top-skill, each with complements or substitutes, and CES). From the systematic pattern of thick tails in the distribution in large cities that we observe below, we can qualitatively deduce that this is due to the complementarities between extreme skills. As in the macro literature on differential complementarities (most notably Krusell, Ohanian, Violante and Rios-Rull, 2000), we obtain information about the technology from the observed equilibrium distribution. We believe that the theoretical link between the complementarities and the distribution is novel. Moreover, from a theoretical viewpoint, we are able to explicitly analyze a tractable matching problem à la Kelso and Crawford (1982) with complementarities (peer effects) that are applied in a concrete labor market setting. Thus far, only general properties such as existence rather than explicit characterizations have been analyzed in these models.

We have chosen to model city difference by means of exogenously given TFP differences. In reality, there are reasons why the productivity of cities is endogenous. We report a model with endogenous agglomeration externalities in the Additional Material Section, available online. The main finding is that agglomeration externalities can lead to asymmetric equilibria with cities of different sizes, even if they are ex ante identical. This occurs provided the external effect is strong enough. We further show that once cities are heterogeneous, the thick tail results extend to this setting with endogenous externalities. And in addition to these production externalities, there could be consumption externalities from the presence of amenities. Unfortunately in our analysis, because we identify unobservable skills from wages, we cannot jointly determine skills and amenities from the same wages. While there is no doubt that amenities matter for citizens' location decisions, based on evidence from Albouy (2008) there seems to be no systematic relation to city size, unlike the relation of skill composition to city size as we derive in our results.

In Appendix C, we also discuss an alternative explanation, namely that thick tails are generated by the combination of top-skill complementarity and the preference for services in a model with home production and a market for services. The intuition behind this model generating thick tails is simple: Top-skill complementarity would attract highly skilled workers to large cities, while the demand for services generated by these highly paid workers would attract low-skill workers to large cities as well. However, as we show in the appendix, under reasonable parameter values, this set up would only generate thick tails under very specific conditions. First, the income share of services must be almost as high as the share on housing expenses. Second, the entry cost in the service sector must be sufficiently high. Finally, the top-skill complementarity must be not too strong. These conditions are very specific and it is not at all clear that they are supported by the facts.

Finally, we also report some further results on housing and consumption 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 1 *Consider a general technology F. 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: $p_j h_{ij} = \alpha w_{ij}$. Since $w_{i1} > w_{i2}$, 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 $h_{ij} = \alpha w_{ij}/p_j$. Since:

$$\frac{w_{i1}}{p_1} < \frac{w_{i2}}{p_2} \tag{14}$$

it follows that $h_{i1} < h_{i2}$. ■

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

Corollary 4 *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 housing prices increase as well, there will be substitution away from housing to the consumption good. As a result, inequality in consumption expenditure will increase.

5 The Empirical Evidence of Thick Tails

We use the one-to-one relation between skills and equilibrium utility to back out the skill distribution from observable variables. The worker's indirect utility in equilibrium is independent of the city, given perfect mobility, and assuming Cobb-Douglas preferences, it satisfies

$$U_i = \alpha^{\alpha} (1 - \alpha)^{1-\alpha} \frac{w_{ij}}{p_j^{\alpha}} \tag{15}$$

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

5.1 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 3 in the appendix for the list of the largest and smallest cities and their 2009 population.

We use wage data from the Current Population Survey (CPS) for the year 2009. We observe weekly pre-tax earnings for 76,821 full-time workers in 254 U.S. metropolitan areas. CPS wages are top-coded

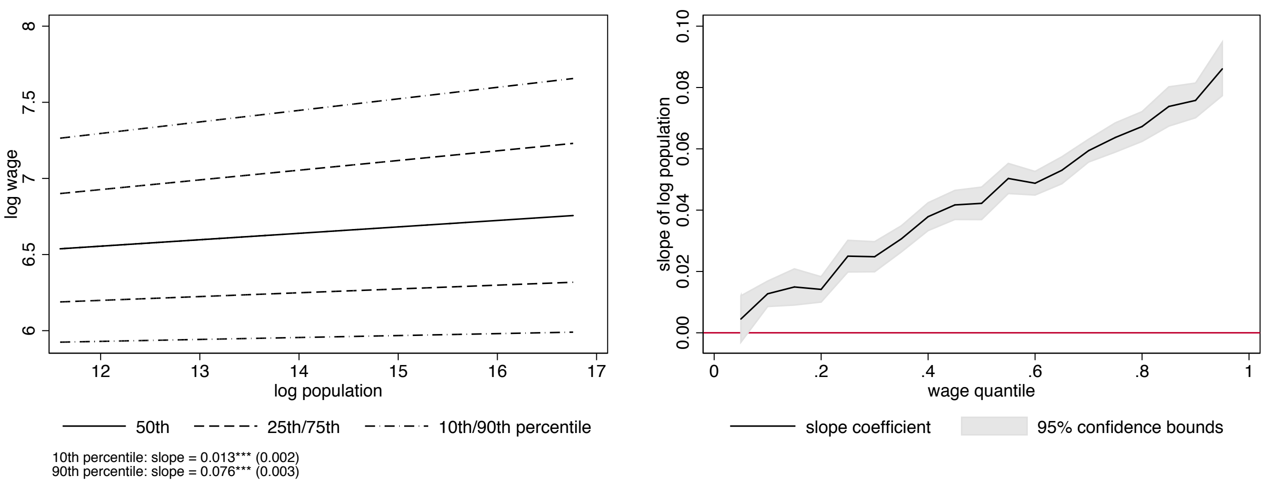

Figure 2: Quantile regression of wage on population. A. 5 selected quantiles; B. Estimated slope for all quantiles.

at around $150,000 which we will take into account in the statistical analysis. In the data appendix, we provide detailed information on data source, sample restrictions and variables.

Local housing price levels are estimated using the American Community Survey (ACS) for 2009. We observe monthly rents for 273,761 housing units in 533 CBSAs. The ACS reports the number of rooms, the age of the structure, and the number of units in the structure. With these data we estimate city specific housing price indices using hedonic regressions. See the appendix for details and a theoretical motivation of this approach.[8]

5.2 Wage Distribution

Figure 1.A in the introduction shows the distribution of weekly wages for full-time earners both in cities with a population of more than 2.5 million and in cities with a 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 fatter in large cities.[9] A simple t-test shows that wages in large cities are 13.3% higher than in small ones ($t = 27.8$, $p < 0.01$). Controlling for right censoring from top-coding and weights in a censored (tobit) regression leads to almost exactly the same comparison: Δ log wage $= 13.2\%$ (robust $t = 24.7$, $p < 0.01$). A look at the tails of the two distributions shows that the large cities have a thicker tail at the top and the small cities at the bottom. The 90th percentile for large cities is 7.56 compared

[8] In an earlier version of this paper we show that our findings are robust to using other housing price data such as from the 2000 U.S. Census, the National Association of Realtors or the Council for Community and Economic Research (C2ER).

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

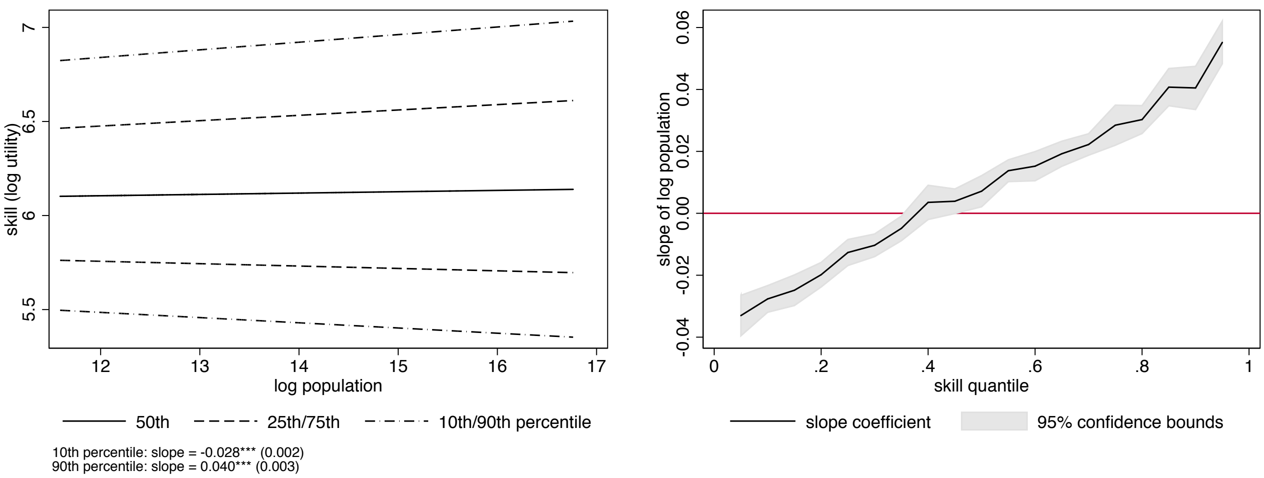

Figure 3: Quantile regression of skills (utility) on population. A. 5 selected quantiles; B. Estimated slope for all quantiles.

to 7.36 for small cities ($\Delta = 0.198$, $se = 0.007$, $p < 0.01$). The 10th percentile for large cities is 5.99 compared to 5.93 for small cities ($\Delta = 0.065$, $se = 0.007$, $p < 0.01$).[10]

The above partitioning of wages into a group of small cities and a group of large cities is arbitrary. We therefore perform quantile regressions of wages on city population size. Panel A in Figure 2 shows the estimated regression lines for the 10th, the 25th, the 50th, the 75th and the 90th percentile. Panel B in Figure 2 shows the slope coefficients for all quantiles. The slopes are all significantly above 0, which implies that the upper tail of the wage distribution increases with city size while the lower tail decreases. For the median (50th percentile), for example, the slope is 0.042 ($se = 0.002$, $p < 0.01$): a doubling of city size leads to a 4.2% increase in wages.

5.3 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 (15).

The variation in housing prices is substantial. While wages increase by 4.2% as city size doubles, housing prices increase by 16.9% for the same change in city size, i.e. a fourfold increase. With the 0.24 expenditure share, this implies that the average cost of living is of a factor $1.169^{0.24} = 1.038$. In

[10] Percentiles and their difference are estimated in a quantile regression of wages on a dummy variable for large cities. We use CPS earnings weights and bootstrapped robust standard errors.

other words, the 4.2% wage gain from living in a larger city is virtually completely absorbed by a 3.8% disutility increase due to the cost of living.

Figure 1.B in the introduction shows the *entire* distribution of skills for full-time earners both in cities with a population of more than 2.5 million and cities with a 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 consistency with the theoretical prediction of thick tails from extreme-skill complementarity.[11] An explicit look at the tails of the two distributions confirms the the thick tail prediction in a statistical sense. The 90th percentile for large cities is 6.99 compared to 6.86 for small cities ($\Delta = 0.132, se = 0.009$, $p < 0.01$). The 10th percentile for large cities is 5.36 compared to 5.44 for small cities ($\Delta = -0.074, se = 0.006$, $p < 0.01$). So the large cities have both a significantly lower 10th percentile and a significantly larger 90th percentile, which implies the thick tails.

As with the wage distribution, one could argue that our partitions of cities into small and large ones is arbitrary. We therefore also run quantile regressions of our implicit skill measure on city population. Figure 3 visualizes the results of these regressions. It shows that the median (50th percentile) barely changes with city sizes while the lower percentiles significantly decrease and the upper percentiles significantly increase. This reiterates our finding that the average of skills does not change systematically with city size while the variance of skills increases significantly. The quantile regressions also perfectly account for the top coding in the wage data up to about the 95th percentile.

6 Additional Sources of Heterogeneity

In this section we allow for additional heterogeneity in locations, in individual preferences, and in prices. We show that our main theoretical results and empirical findings are robust to allowing for heterogeneity in attractiveness of locations within cities (section 6.1), for non-homothetic household preferences (section 6.2), and for local price variation of other consumer goods beyond housing (section 6.3).

6.1 Heterogenous Attractiveness of Locations Within Cities

In our analysis, the endogenous choice of housing is a central component. High-skill workers in the same city consume more housing h than low-skill workers, and at the same time there is substitution

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

between housing and other consumption goods: same skilled workers consume less housing in large, expensive cities than in small ones. In order to obtain our wage-based measure of skills, and guided by the theory, we adjust wages by a city-wide housing price index, which measures the unit cost (say, per square foot) of housing. To adjust for different choices of quantities, we have obtained that unit cost by means of a hedonic regression that conditions on observables, such as the number of rooms, bathrooms, etc. Implicit in this specification is the assumption that all neighborhoods are equally attractive, and citizens with different incomes will therefore share the same neighborhoods, albeit in houses of different sizes.

In reality though, not all locations within a city are equally attractive. The desirability of neighborhoods depends on factors such as closeness to work, cultural events, restaurants, shopping outlets and recreational opportunities, as well as on access to good schools and other public goods, or simply on its socio-economic status. In spatial equilibrium, more attractive locations will have higher housing prices. In the absence of within-city sorting – which will be discussed further down – all households are indifferent between all locations within the city in spatial equilibrium. A simple theoretical framework of this mechanism is the mono-centric city model where the attractiveness of locations within the city decreases with distance from the central business district due to commuting costs.[12] Because there is a tradeoff, the housing price now also reflects the attractiveness. The less attractive, the lower the housing price, even for identical agents. As a result, there is a so-called bid-rent function that increases with attractiveness. Only in the city center without any commuting, does the price reflect the true cost of living. Anywhere else, the housing price is too low since it embodies both the cost of living and the disutility from less local attractiveness.

We operationalize neighborhoods within cities in the data as Public Use Microdata Areas (PUMA), relatively small areas of around 100,000 inhabitants. We estimate hedonic price indices for all PUMA areas across the U.S. We then take the average price index of the top 10% PUMA areas per city, i.e. CBSA. The imputed skill distribution based on this maximal housing price index is reported in Figure 4. Consistent with our earlier results, the thick tails continue to exist. We tend to see a somewhat bigger tail at the bottom than at the top. Even adjusting for differential attractiveness of locations within the city, the thick tails result continues to hold.

So far, the logic with differential attractiveness of locations is for identical agents. In fact, housing prices adjust to equalize the difference in attractiveness, and all agents are indifferent where to locate within the city. When in addition, agents are heterogeneous – as in our model in incomes – and different

[12] For an overview of the different variants of the mono-centric city model and a full characterization, see Fujita (1989).

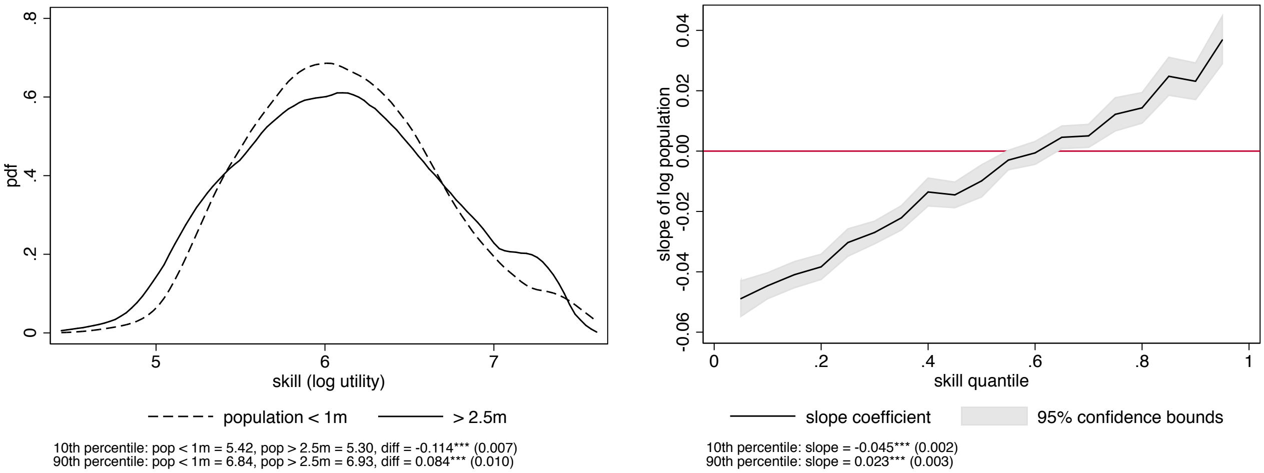

Figure 4: Skills based on the average housing price index in the top 10% neighborhoods (PUMA) of a city; wage data from the 2009 Current Population Survey (CPS). A. Skill distribution for small and large cities. B. Slopes in quantile regressions of log utility on log population.

income agents value the tradeoff between local attractiveness and housing price differently, then there will be *sorting within the city*. If the opportunity cost of attractiveness is complementary with income, then the richest citizens will sort into the most attractive neighborhoods (e.g. Manhattan in New York or Bloomfield Hills in Detroit). Now there is a bid-rent function for each household type that is only observed where a citizen type actually lives: at the most attractive location for the highest income types, in intermediate attractive locations for the middle income types and at the least attractive locations for the lowest income types. In the case of perfect sorting, each type would live in a dedicated neighborhood between intersections of the bid-rent functions. The relevant price for utility comparisons across cities would be the bid-rent at the most attractive location, but that is unobserved except for the highest income types in the most attractive locations. The observed local housing prices for less attractive neighborhoods is then a lower bound of the relevant price since the observed price incorporates the cost of commuting.

We estimate this lower bound as the hedonic price index of the PUMA area where the observed worker lives. Unfortunately, the wage data from the CPS do not identify the PUMA area of the worker. We therefore use wages from the 2009 American Community Survey (ACS) for this analysis, i.e. wage data from the same source as the price data. We then assign each worker in the ACS the housing price of the PUMA area where he lives. In Figure 5, we first reproduce the basic findings from Figure 1. As with CPS data, the ACS data shows first order stochastic dominance in wages and thick tails in skills,

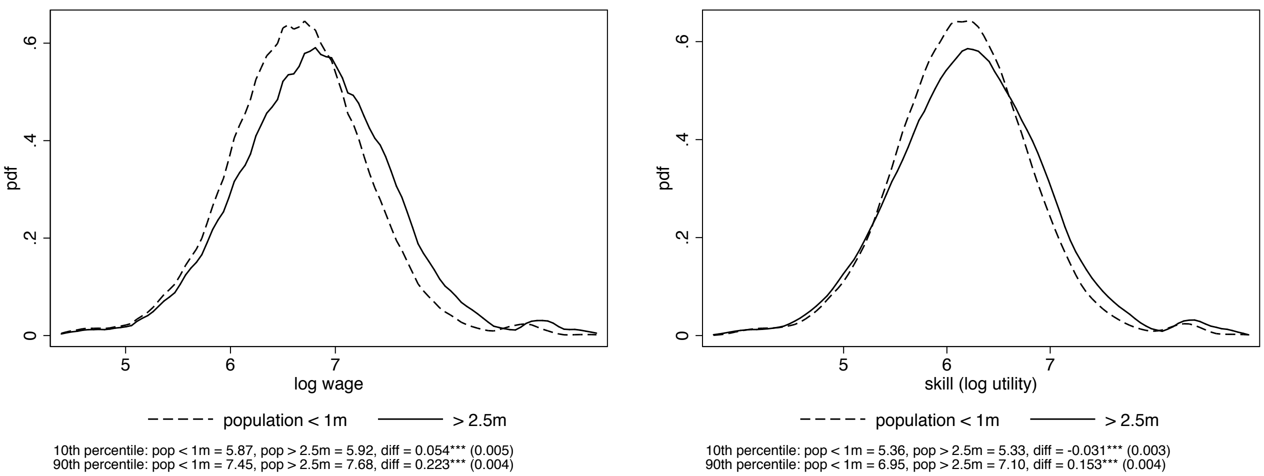

Figure 5: Wage and skill distribution for small and large cities; wage data from the 2009 American Community Survey (ACS); Kernel density estimates (Epanechnikov kernel, bandwidth = 0.1); A. Wages; B. Skills.

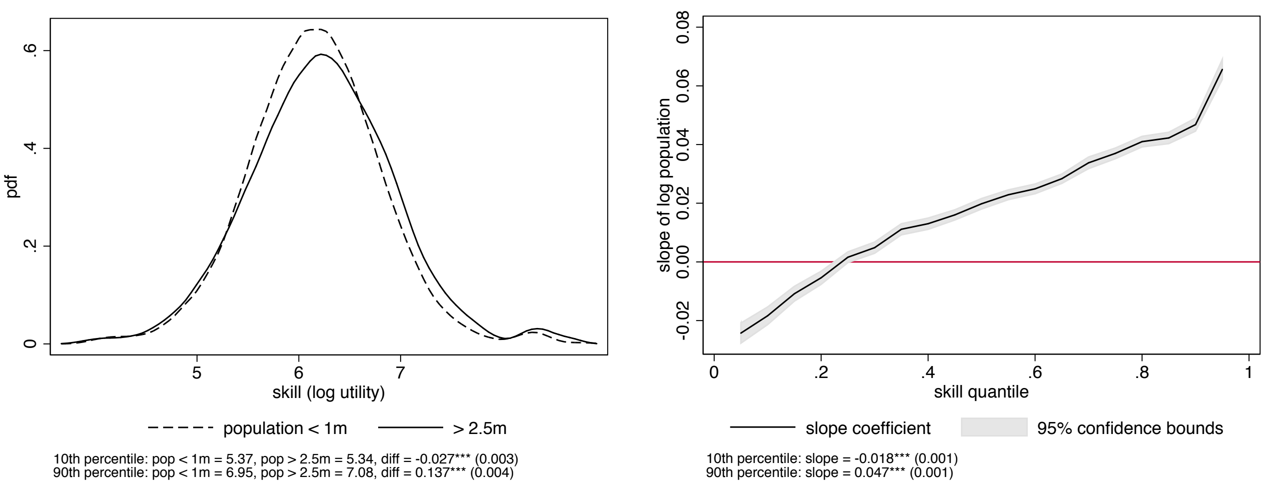

Figure 6: Skills based on the housing price in the neighborhood of residence; wage data from the 2009 American Community Survey (ACS). A. Skill distribution for small and large cities. B. Slopes in quantile regressions of log utility on log population.

though the effect on the lower tail is less pronounced.[13]

In Figure 6, we report the distribution of skills based on the price of the PUMA area where the worker lives. Quite remarkably, even in the presence of this biased price index, the thick tails continue to exist. It is not surprising that the lower tail difference is thinner, given that our measure is biased downwards, but it is still significant. More importantly, because this housing price index is a lower

[13] The CPS is generally considered the more reliable data source for wage data as the survey is performed personally by phone while the ACS questionnaires are mailed. We therefore use our initial results in Figure 1 as the baseline.

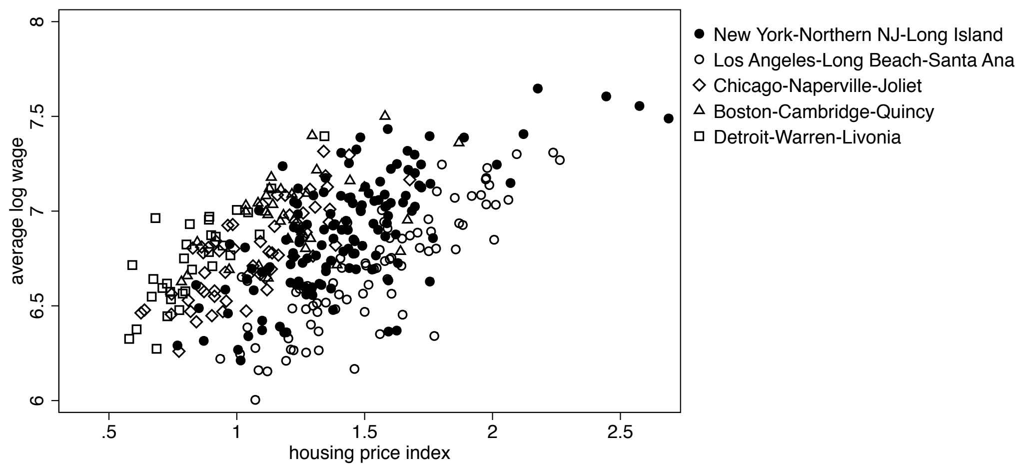

Figure 7: Average rental housing prices and average log wages across PUMA areas in 5 selected CBSAs.

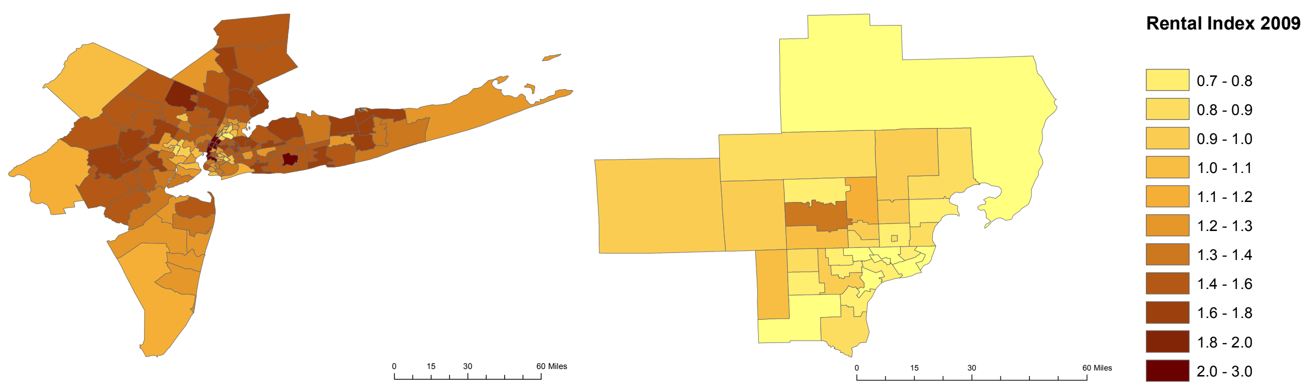

Figure 8: Average rental housing prices across PUMA areas in A. New York-Northern New Jersey-Long Island CBSA and B. Detroit-Warren-Livonia CBSA . The housing price index is normalized to 1 for the US sample average.

bound, the actual tail difference must be thicker.

We interpret the exercise in this section as one in which we put bounds on the tails. The neighborhood (PUMA) price index is the lower bound and shifts the distribution too little to the left, resulting in a small lower tail effect. The maximum price index is the upper bound and shifts the distribution too much to the left, generating a big lower tail effect but hardly any upper tail effect. The relevant price index, and hence the distributions, is somewhere in between.

We conclude that allowing for differential attractiveness and ensueing within-city sorting, the thick tail result continues to hold. That does not mean that there is no within city sorting, quite to the contrary. Figure 7 shows the average housing prices and average log wages across neighborhoods (PUMA areas) for 5 metro areas. The fact that there is a strong correlation between wages and housing prices is indicative of such within city sorting.

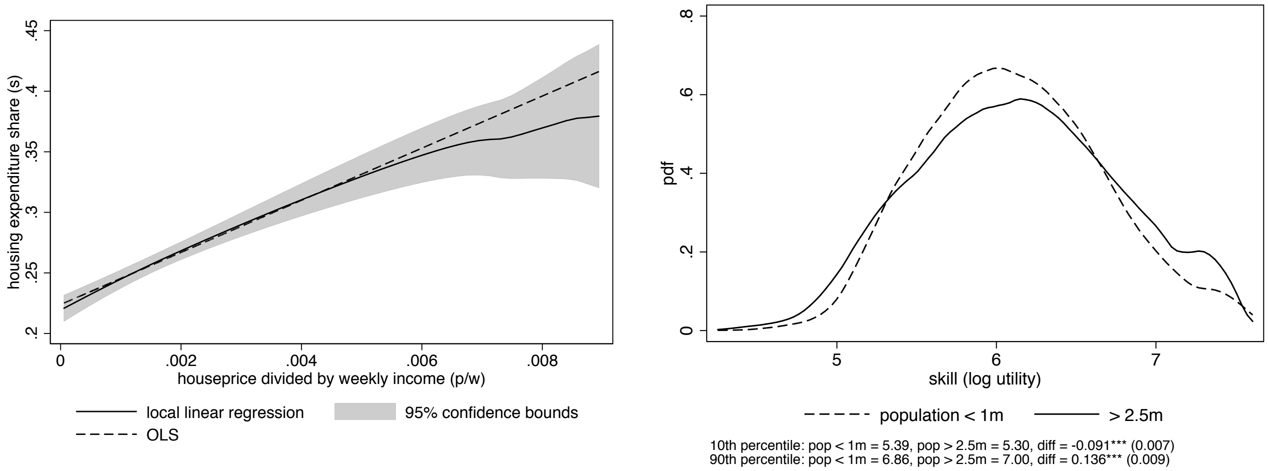

Figure 9: Expenditure Share (A) and Skill distribution (B) using Stone-Geary preferences, $\widehat{\alpha} = 0.224$, $\widehat{\underline{h}} = 27.7$.

A closer look at the data reveals that the price difference across different neighborhoods within a city is smaller than the price variation across cities: the standard deviation of the puma housing price index is 0.36 across and within all CBSAs; the standard deviation *between* CBSAs is 0.33 while it is only 0.25 *within* CBSAs. This is illustrated in Figure 8 with the price variation across PUMA areas in two cities, New York and Detroit. Among the most expensive residential areas in New York, for example, are the Upper East Side in Manhattan and East Meadow on Long Island. The least expensive areas are East Harlem in Manhattan and the Bronx. Yet, few of the cheapest areas in New York are less expensive (less dark on the map) than the most expensive areas in Detroit. The important point to take away from this is that the housing price differences across cities is relatively large.

6.2 Non-homothetic Household Preferences

So far we have assumed homothetic (Cobb-Douglas) preferences over housing and consumption. Motivated by the empirical finding of Davis and Ortalo-Magné (2009) that housing expenditure is on average remarkably constant across different cities, we have used their estimated expenditure share on housing of $\hat{\alpha} = 0.24$. Yet, even if the average expenditure share of housing is constant across cities of different size, there may well be variation across individuals of different incomes. As a result, the Engel curve that relates expenditure to income is no longer linear as it is under Cobb-Douglas. Below we show that there is indeed evidence in our data of a concave Engel curve: the rich spend proportionally less on housing.

Non-homothetic preferences have important consequences for both our theoretical model and our empirical strategy. First, decreasing housing expenditure shares with respect to income introduce an alternative mechanism for sorting across cities as high-skill workers care less about local housing prices than low-skill workers.[14] Second, our price-based skill measure derived from the homothetic Cobb-Douglas preferences needs to be adjusted.

It is straightforward to introduce non-homothetic preferences into our theoretical framework. We follow the most common way in the literature and model it by means of Stone-Geary preferences. They can be written as $u(c,h) = c^{1-\alpha}(h-\underline{h})^{\alpha}$ where $\underline{h}$ is the subsistence level of housing, and housing consumption is restricted to $h \geq \underline{h}$. Given housing prices p and the budget constraint $c + ph \leq w$, from the first order conditions optimal expenditures on housing and consumption can be written as $ph^\star = \alpha w + (1-\alpha)p\underline{h}$ and $c^* = (1-\alpha)(w - p\underline{h})$, with the indirect utility given by

$$u(c^\star, h^\star) = (1-\alpha)^{1-\alpha}\alpha^{\alpha}p^{1-\alpha}\left(\frac{w}{p} - \underline{h}\right). \tag{16}$$

Assuming the CES production technology and Stone-Geary preferences with $\underline{h} > 0$, our model predicts FOSD of the skill distribution in large cities in simulations.[15] Hence, non-homothetic preferences do not generate thick tails per se. But allowing for extreme-skill complementarities in addition, our model still predicts thick tails, for $\underline{h}$ positive but small. In the next paragraph we therefore explore which mechanism prevails empirically.

When housing expenditure varies by income, the utility and therefore our measure of skill must be adjusted. Assuming Stone-Geary preferences, the expenditure share on housing is a linear function in the *inverse* of wages: $\frac{ph^\star}{w} = \alpha + (1-\alpha)\underline{h}\frac{p}{w}$. This give us a regression of the housing expenditure share s_i on p_j/w_i:

$$s_i = \alpha + \beta\frac{p_j}{w_i} + \varepsilon_i \tag{17}$$

where $s_i = \frac{p_j h_i^\star}{w_i}$. The parameter α is estimated directly while the parameter $\underline{h}$ is estimated as $\hat{\underline{h}} = \hat{\beta}/(1-\hat{\alpha})$.

We use individual data on expenditure shares (see the data appendix for details) from the Consumer Expenditure Survey (CEX) to estimate the two parameters α and $\underline{h}$. We obtain $\widehat{\alpha} = 0.224$ (s.e.= 0.005), and $\widehat{\underline{h}} = 27.7$ (3.8). The implied expenditures shares vary considerably, from 35% for low-income households to 22% for high-income households as graphed in Figure 9.A. Figure 9.A. also shows that the functional form assumed by Stone-Geary fits the data astonishingly well. Yet, the varying expenditure share and the resulting non-linearity of the Engel curve do not substantially alter the

[14] See e.g. Schmidheiny (2006) who studies within-city sorting from assuming non-homothetic (Stone-Geary) preferences.

[15] Matlab code for the simulations can be obtained on request.

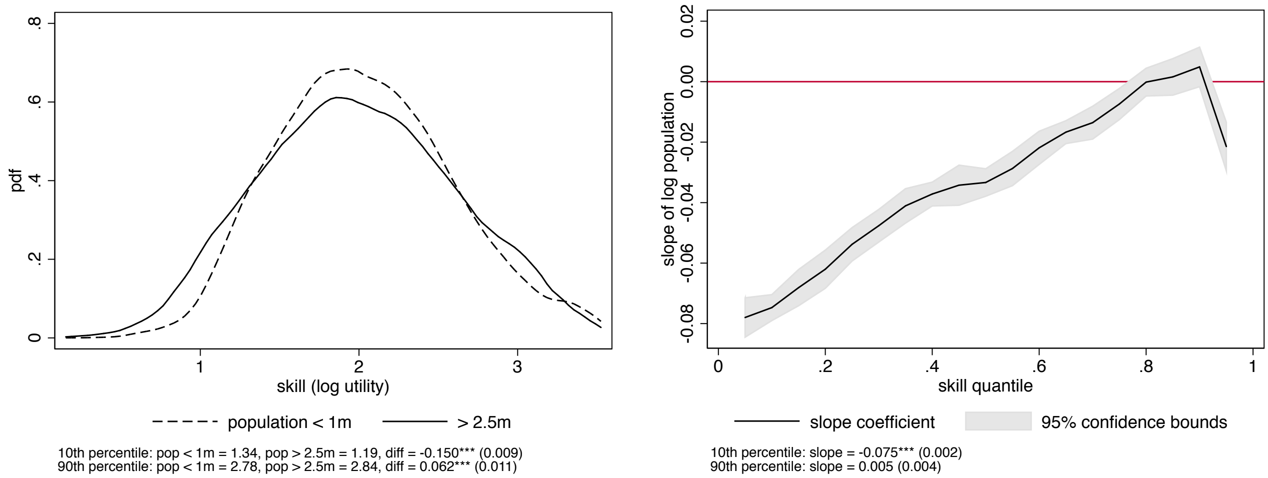

Figure 10: Skill distribution using ACCRA cost-of-living index, adjusting for variation in prices for all goods. A. Skill distribution for small and large cities. B. Slopes in quantile regressions of log utility on log population.

emergence of thick tails. Figure 9.B. shows the resulting skill distribution, which has very similar properties to those under Cobb-Douglas preferences. If anything, the thick tail differences in both the lower and the upper tail are slightly more pronounced. This shows that our evidence for extreme-skill complementarities still holds even after accounting for non-homothetic preferences.

6.3 Variation in Consumption Prices

Local prices are crucial for our strategy to back out skills from observed nominal wages. In this section we look not only at local housing prices but also at local prices of consumption goods. It may well be that consumption prices in large cities are systematically higher than in smaller cities, thus adding 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). See the data appendix for details. 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 9.1 for grocery items, 14.2 for utilities, 5.6 for transport, 8.2 for health and 4.6 for services; all price indices are normalized to mean 100).

Figure 10 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.[16] When including the price index for all consumption and housing,

[16]ACCRA reports a composite price index which is the weighted average of the six sub-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 $P_{composite} = (P_{grocery})^{\alpha_{grocery}} \cdot ... \cdot (P_{services})^{\alpha_{services}}$. The implied skills are calculated as $U_i = w_{ij} / (P_{grocery,j}{}^{0.1249} \cdot P_{housing,j}{}^{0.2918} \cdot$

we find that the left tail difference becomes more pronounced while the right tail difference is 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 existence of fat tails. Note again, that the the third crossing at the very top is an artefact of the top-coding (see footnote 11).

These findings should be interpreted with some caution and a few caveats. First, the quality of the ACCRA data is dubious.[17] 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 pays higher prices, a price index incorporating the diversity and quality on offer will be lower.[18] In addition, ACCRA puts more weight on housing than we do in our baseline result (see details in footnote 16). Using our budget share, the skill distribution of large cities would be shifted less to the left, the lower tail difference would be less pronounced and hence more in line with our baseline results. We therefore see the results in Figure 10 as a very conservative upper bound of how the inclusion of consumption price differentials affects our initial findings.

7 Decomposing Observable and Unobservable Skills

Our measure of skills is a price based measure, calculated as the residual of wages after adjusting for housing prices. As such, it is a comprehensive measure that encompasses both observed and unobserved characteristics. In this section we verify the robustness of the spatial sorting result using observed, direct measures of skill. Not only does it allow us to decompose the contribution of observed and unobserved characteristics, it also makes explicit if and why the results with our comprehensive measure are different from a long tradition of previous literature. We show that our thick tail result also holds using directly

${P_{utilities,j}}^{0.0998} \cdot {P_{transport,j}}^{0.111} \cdot {P_{health,j}}^{0.0406} \cdot {P_{services,j}}^{0.3319}$). The exponents are budget shares taken from the ACCRA Cost of Living Index Manual, version November 2009 (current versions of the manual are available online at http://www.coli.org/surveyforms/colimanual.pdf). The ACCRA values differ from the ones we use for the baseline results of this paper. Based on Davis and Ortalo-Magné (2009), we use a budget share of 0.24 for housing rather than the 0.29 used by ACCRA.

[17]Koo et al. (2000) discuss several problems of the ACCRA data.

[18]This also appears to be an issue when studying price differences across different countries. Comparing 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.

observed measures of skill. We also contribute to the debate on the role of non-cognitive skills and how much they contribute to wage determination. Here, we focus explicitly on the impact of non-cognitive skills across labor markets of different size.

In the first instance (section 7.1), we use years of schooling, occupation and industry as direct measures of skills instead of our wage based measure. We then investigate the role of mobility and migration that is systematic by nationality (7.2) and also whether there are any location decisions determined by age over the life cycle (7.3). Finally, we quantify the impact of those observables simultaneously (7.4) and decompose the tails into an explained and an unexplained component.

7.1 Observed Measures of Skills: Education, Occupation and Industry

As a first robustness check and as external validation, we compare our implicit skill distribution with that of educational attainment. The top-left panel in Figure 11 shows the distribution of highest educational attainment for the same CPS population as our wage data, where workers are grouped in 5 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 (population above 2.5M) than in smaller ones (population between 100,000 and 1M). What is most striking about this observation is that the thick tails in the distribution of educational attainment are obtained *independently* of how we constructed our measure of skills before. Here, no theory is needed and the measure of skills is determined exogenously. Using observable, self-reported measures of skills we find a distribution with thicker tails in larger cities, both in the aggregate and at the individual city level.

In principle, individual skills can be decomposed into an observed component, e.g. education, and an unobserved component, e.g. ability. We already know that the *observed* component indeed exhibits thick tails. To get to the unobserved component, we regress our implicit skill measure (log utility) on dummy variables for all 16 observed education categories.[19] The residual of this regression is the "residual skill" after controlling for observed education. A high value means that the worker is more skilled relative to other workers with the same education. This can be a very successful lawyer or a high-school-dropout-become-succesful-entrepreneur.

[19] We estimate censored (tobit) regression, accounting for the top-coding of the wage data. We regress $\log(u_{ij})$ on a constant and a set of dummy variables for education with basic education as reference group. This dummy variable regression is fully consistent with our theoretical model. Recall that the wage ratio of skill type i relative to skill type 1 is constant across cities, $\log(w_{ij}/w_{1j}) = \log(w_{ij}) - \log(w_{1j}) = \beta_i$ and therefore the ratio of log utility, too, $\log(u_{ij}/u_{1j}) = \log((w_{ij}/p_j^{\alpha})/(w_{1j}/p_j^{\alpha})) = \log(w_{ij}/w_{1j}) = \beta_i$. The log utility of skill type i can therefore be expressed as $\log(u_{ij}) = \beta_1 + \beta_2 * d_2 + ... + \beta_i * d_i + ...$ where β_1 is $\log(u_{ij}) = \log(u_{i1}) = \log(u_i)$ of the reference skill type 1, which is constant across cities j. Notice, that regressing log(wage) on a constant and dummy variables for education would not be consistent with our theoretical model as the constant, $\log(w_{1j})$, would be city specific. The usual wage regression therefore needs city fixed effects, which we do not need.

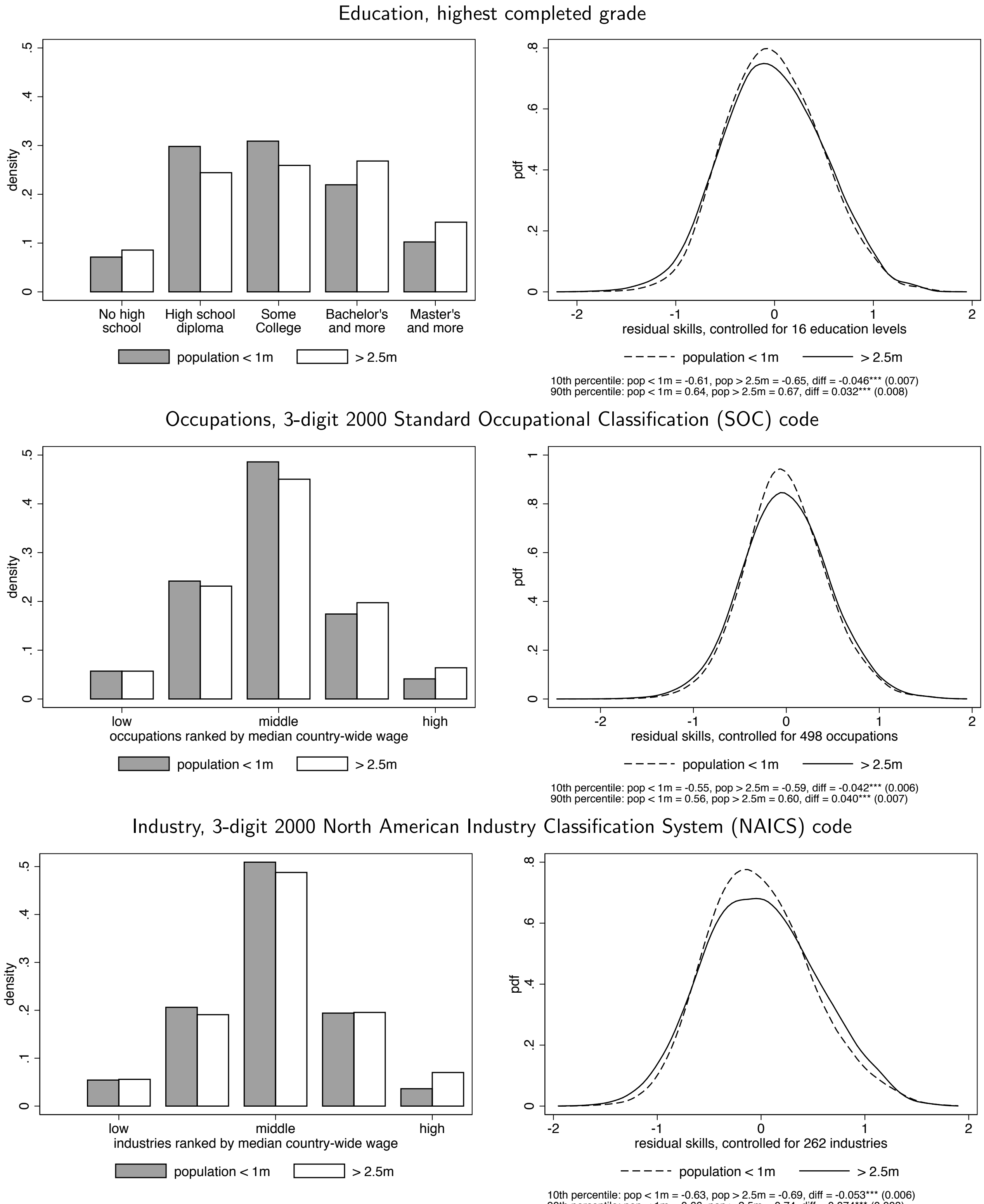

Figure 11: Observed and residual skills of full-time wage earners in 2009 CPS for small and large cities.

The top-right panel in Figure 11 shows the resulting distribution of residual skills for large and small cities. We see that the thick tails persist for the unobserved component, and are more pronounced at the bottom than at the top. Large cities seem to host more of the relatively *low* skilled even after controlling for education category. Taken together, these two pieces of evidence mean that large cities attract relatively more of the best educated and more of the least educated workers. In addition, they attract the more talented, given their education level. We also see that there is an asymmetry between the upper and the lower tail. This indicates that the sorting of the most talented workers is based more on the observed component of skills (education) while the sorting of the least talented is based more on the unobserved component. A finding that we will reiterate in the next paragraph, where we condition on occupations.

Like education, occupation can be interpreted as a direct measure of skills. More highly skilled individuals are likely to be employed in higher ranked occupations. If we can find a way to rank occupations, it can give us some insights into the sorting of workers based on this observed characteristic. The current population survey contains the occupation of workers, in addition to wages and education. This occupational classification has been used as an alternative direct measure of skills before (see e.g., Autor, Levy, Murnane, 2003). The CPS reports occupations in 498 categories from the 2000 Standard Occupational Classification by the U.S. Bureau of Labor Statistics. We use country-wide median wages by occupation category as a proxy of their skill requirement (this approach is used in e.g. Goos, Manning, Salomons, 2009). We then group the 498 occupation categories into 5 groups: group 1 includes the lowest paid occupations covering 5% of all workers, group 2 the next 20%, the middle group 3 covers 50% of workers, group 4 the next 20%, and the high skill group 5 includes the highest paid occupations covering 5% of all workers. The low skill group 1 includes occupations such as dishwashers, waiters, and child care workers; the middle group 3 occupations like secretaries and truck drivers; the high skill group includes chief executives, surgeons and lawyers. The middle-left panel in Figure 11 shows the distribution of these 5 occupation groups separately for small cities and for large cities. It shows that workers in the highest paid occupations locate relatively more often to large cities while the middle occupations locate more to small cities. The lowest paid occupations are equally frequent in small and large cities. Note, that already the fact that the lowest group is *not more* frequent in small cities is a contradiction of the first order stochastic dominance observed in nominal wages. This direct evidence of thick tails is very similar to the result we found for education groups. The effect on the lower tail is less pronounced than for the upper tail.

As with the analysis of education, we next decompose skills into a component observed through occupations and the residual unobserved component. We regress our skill measure (log utility) on dummy variables for all 498 occupation categories.[20] The residual of this regression is the "residual skill" after controlling for observed occupation. The middle-right panel in Figure 11 shows the resulting distribution of residual skills for large and small cities. We see that the thick tails persist for the unobserved component. In particular, large cities seems to host more of the relatively low-skilled in each occupation than small cities. As a result, large cities attract relatively more workers in the highest paid occupations and more in the lowest paid occupations. In addition, they attract the more talented, given their occupation. We also see that the sorting of the most talented workers is based more on the observed component of skills (occupation) while the sorting of the least talented is based more on the unobserved component. A finding that we already made using education as a direct measure.

A potential threat to our identification of skills is the different industrial composition of cities. Wages systematically differ across industries and the mobility of workers varies across sectors (see for example Davis, Faberman and Haltiwanger, 2006). At the same time, industrial composition changes across cities: some cities specialize in particular industries and other cities have a diverse industrial mix. If industry composition varies systematically with city size, this could be an alternative explanation for our finding of thick tails. We therefore seek to control for the wage component related to the industry. As in the analysis of occupations, we first rank the 262 industries by country-wide median wages and group them into 5 groups: group 1 includes the lowest paid industries covering 5% of all workers, group 2 the next 20%, the middle group 3 covers 50% of workers, group 4 the next 20%, and the high skill group 5 includes the highest paid industries covering 5% of all workers. The middle-left panel in Figure 11 shows the distribution of these 5 industry groups separately for small cities and for large cities. We see indeed that large cities attract more workers from industries that pay the highest wages as well as more workers from industries that pay the lowest wages.

As we did with education and occupation, we decompose our implicit skill measure into a component observed through occupations and the residual unobserved component. We regress our skill measure (log utility) on dummy variables for all 262 4-digit industries.[21] The residual of this regression is the "residual skill" after controlling for the industry the worker operates in. The bottom-right panel in Figure 11 shows the resulting distribution of residual skills for large and small cities. We see substantial thick tails both on the upper and the lower end when controlling for industries. This shows that while the industrial composition may vary across cities, it does not do so systematically across small and large cities.

[20] We estimate censored (tobit) regression accounting for the top-coding of the wage data.

[21] We estimate a censored (tobit) regression, accounting for the top-coding of the wage data.

7.2 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 the foreign born 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 low skilled Americans or low-skilled immigrants who disproportionately locate to 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 to either large or small cities. Thus it may well be the case that migrants have certain benefits from locating to large cities. For example networks (see Munshi, 2003) play an important role in 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.

Consider this logic with preference bias in the context of our baseline model. There are three skill types (low, middle, high) and two cities (small and large). Assume now that all migrants are low skilled and have a preference for the large city, i.e. they get additional utility from living there. Suppose that in an equilibrium without migrant preferences, the number of low skilled workers in the large city is m_{11}. Now with the migrants' preference for large cities, *all* migrants will want to live in the large city. As long as the number of migrants is below m_{11}, the general equilibrium will not change at all. All low skilled migrants will live in the large city, some low skilled natives will also live in the large city and all the others in the small city. The native workers remains the marginal worker and are still indifferent between large and small city. Migrants, however, strictly prefer the large city and are not indifferent. The resulting equilibrium skill distribution is not be affected by migrants. Just looking at migrants, we would see a (very much) thicker lower tail in large cities. Just looking at natives we would see a less thicker (if not thinner) lower tail in big cities. The important implication is that even if we observed those location biases, it would not affect the aggregate predictions.

In general, as long as the migrants – or any subgroup of the population for that matter – of a particular skill and with a taste for a particular city do not outnumber the equilibrium number of workers of that skill in that city, our model predictions are unchanged by migrant preferences for large

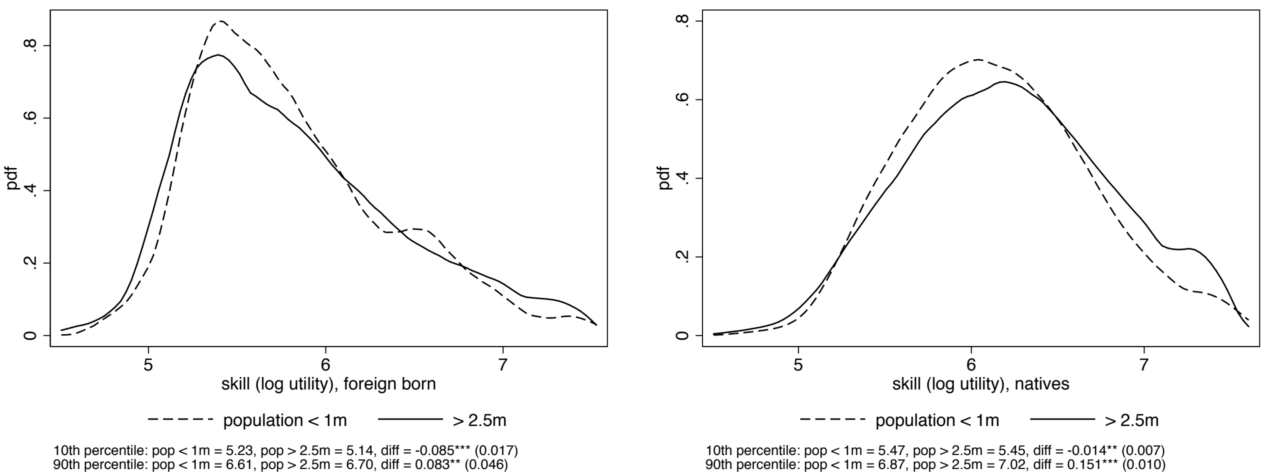

Figure 12: Skill distribution for small and large cities. A. Foreign born workers; B. Natives.

cities. The fact that we observe a substantial number of natives of all skill types (including the very low skilled) in all cities, provides empirical support that the native workers are the indifferent marginal workers for all skill types.

Empirically, we even find no evidence of a systematic bias in the location decision of the migrants. To evaluate this, we split the sample up into natives and foreign born workers. Figure 12 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 also see that the distribution of foreign born workers has thick 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 migrants, the distribution of natives has thicker tails in large cities. The thick tails are therefore not driven by selective migration decisions by non-natives.

7.3 Age: Location Decisions over the Life Cycle

One plausible mechanism, and different from the technological one we propose, is that the spatial sorting pattern is driven by location decisions that vary over the life cycle. We distinguish between three candidate mechanisms in the presence of human capital accumulation. First, age dependent preferences could lead to variation in the location decision over the life cycle. Young people prefer the excitement of the city, while older people settle for a quiet life. Second, family and marriage considerations determine whether to live in an urban or rural environment, for example as in Gautier, Svarer and Teulings (2009). Singles find a better marriage market in the big city, while those married

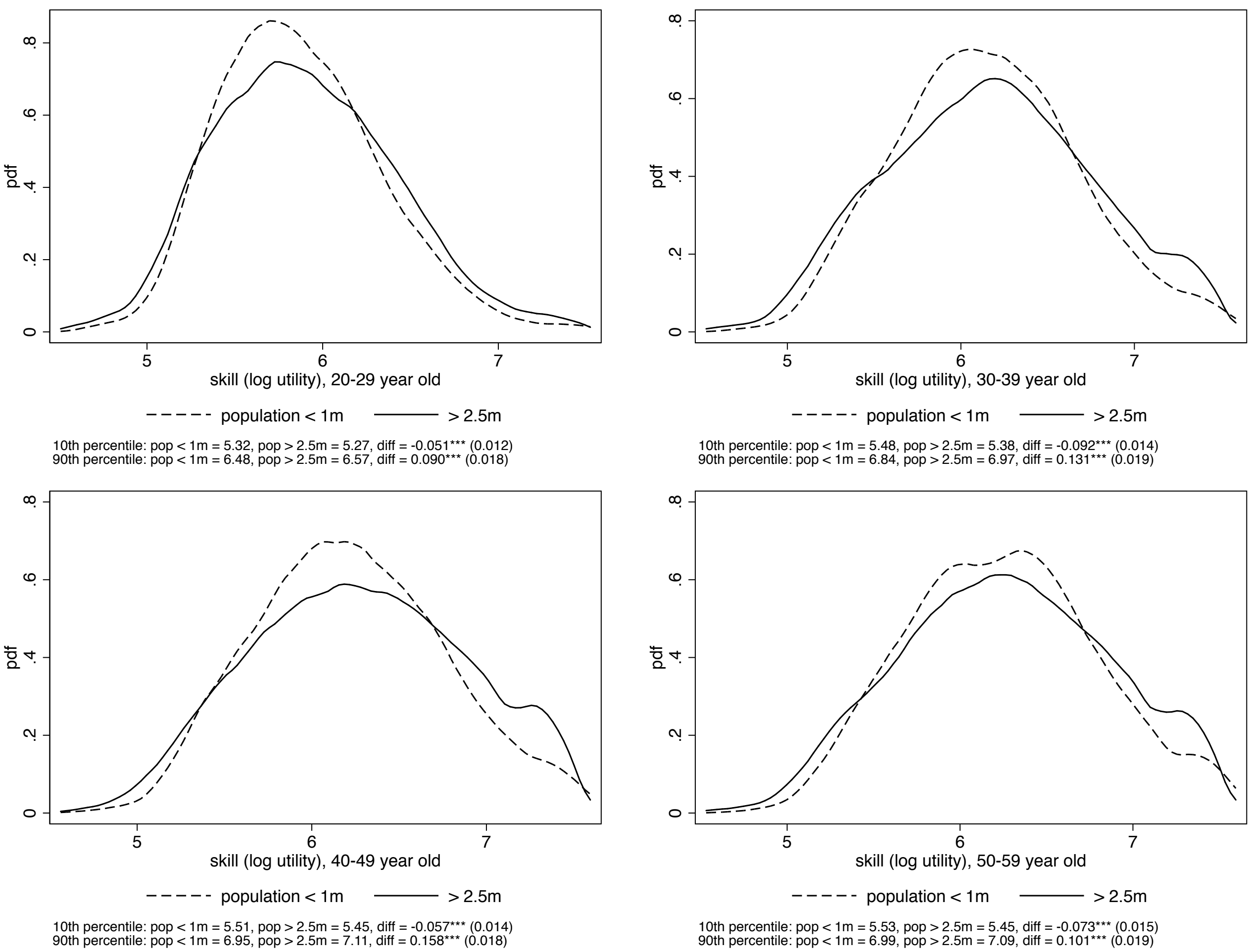

Figure 13: Skill distribution for small and large cities by age group.

with children look for green spaces and schools for the children. Third, labor market learning and human capital accumulation may affect the location decision (as in Puga and De la Roca, 2012). Young workers try their luck in the big city, starting off at low salaries. At a later age, those who have learned (or are lucky) to be very productive stay and earn high salaries, while the unlucky who have learned little and have added limited human capital return to their small town at moderate salaries.

Each of these three mechanism induces a systematic spatial sorting pattern and as a result a systematic skill distribution for the entire cross section. This can lead to thick tails or first order stochastic dominance. Most importantly, if it is life cycle driven, it will differ for different age groups. For example, there may be stochastic dominance of the small cities for the young and stochastic dominance of the large cities for the old, thus leading to thick tails overall. For that purpose, we split the sample into four different age cohorts and investigate the tail properties of the skill distribution for each cohort as reported in Figure 13. For each of the cohorts, there are thick tails, and the statistics show they are highly significant. While we obviously cannot rule out that any of the mechanisms mentioned above is at work, this evidence lends support to the fact that these mechanisms are not sufficiently strong to undo the impact of the extreme-skill complementarities.

7.4 Decomposition by Education, Occupation, Industry and Individual Attributes

This section seeks to explain the differences in the tails of the skill distribution *simultanously* by all three variables in section 7.1 (education, occupation, industry) plus individual attributes (sex, age, race, foreign birth). In regression models explaining the mean, this can easily be done by the Oaxaca-Blinder decomposition (Oaxaca, 1973; Blinder, 1973). However, we are interested in explaining differences in the *tails* of the skill distribution, which is a much harder task. Fortunately, there is a fast evolving econometrics literature on the decomposition of entire (wage) distributions (Juhn, Murphy, Pierce, 1993; DiNardo, Fortin, Lemieux, 1996; see Fortin, Lemieux, Firpo, 2011 for a review).

We are using two very recent approaches. The first approach is based on Chernozhukov, Fernández-Val, Melly (2013). This approach estimates the entire distribution of skills *conditional* on the observed covariates for both groups (small and large cities) separately using quantile regressions. They then integrate the conditional distributions over the covariates to get the predicted *marginal* distribution of each group. With this they can predict counterfactual marginal distributions such as the distribution of skills in the large city, given that it had the same distribution of covariates as the small city. We refer to the difference between the marginal distribution and the counterfactual distribution as the "unexplained" difference.[22]

We summarize the results by reporting the impact on the 10th and 90th percentile. Chernozhukov, Fernández-Val, Melly (2013) cannot easily decompose the explained difference into different sets of covariates. We therefore also apply an alternative decomposition proposed by Firpo, Fortin, Lemieux (2009). Their approach, based on so-called rescaled influence functions (RIF), allows to conveniently decompose the explained differences in the quantiles into the contribution of each covariate. A downside of Firpo, Fortin, Lemieux (2009) is that the basic approach is less intuitive.[23]

The original classification into 16 education categories, 498 occupations and 262 4-digit industries

[22] In labor economics, the unexplained difference is often called a "wage structure effect". This is because the difference in the conditional distributions between the two groups may stem from different wage schemes, i.e. returns to e.g. education. However, it may also stem from different conditional (i.e. residual) skill distributions in the two groups. The decomposition by Chernozhukov et al. and Firpo et al. do not and cannot disentangle wage structure from residual skills. We do not take a stance at either interpretation and simply call it the unexplained difference.

[23] Firpo, Fortin, Lemieux (2009) is an approximation for a *marginal* location shift of the distribution of the covariate. It is not known how good this approximation is if the change in the distribution of the covariate of interest is large or if the covariate of interest is discrete (like dummy variables). See Rothe (2012, Appendix B) for a discussion.

Table 1: Decomposing the skill distributions of large and small cities.

	10% Quantile			90% Quantile		
Observed Quantiles:						
- Large cities	5.365	(0.004)	***	6.994	(0.006)	***
- Small cities	5.439	(0.005)	***	6.862	(0.007)	***
- Difference	-0.074	(0.006)	***	0.132	(0.009)	***
Firpo, Fortin, Lemieux (2009)						
Predicted Quantiles:						
- Large cities	5.387	(0.005)	***	7.022	(0.005)	***
- Small cities	5.454	(0.004)	***	6.878	(0.008)	***
- Difference	-0.068	(0.007)	***	0.144	(0.009)	***
Explained by observables:						
- Education (16 categories)	0.003	(0.002)	**	0.052	(0.002)	***
- Occupation (22 categories)	0.004	(0.002)	*	0.025	(0.003)	***
- Industry (51 categories)	-0.001	(0.002)		0.013	(0.002)	***
- Race (4 groups)	-0.004	(0.001)	***	-0.015	(0.001)	***
- Sex	-0.001	(0.001)	*	-0.002	(0.001)	*
- Foreign born	-0.020	(0.002)	***	-0.004	(0.001)	***
- Age (2nd order polynomial)	0.000	(0.001)		-0.002	(0.001)	*
Total explained by observables	-0.018	(0.004)	***	0.067	(0.005)	***
Not explained by observables	-0.049	(0.006)	***	0.077	(0.008)	***
Chernozhukov, Fernández-Val, Melly (2013)						
Predicted Quantile difference	-0.068	(0.006)	***	0.113	(0.009)	***
Explained by observables	-0.019	(0.004)	***	0.064	(0.005)	***
Not explained by observables	-0.050	(0.007)	***	0.049	(0.007)	***

Notes: Large cities: population > 2.5m; small cities: population <1m. Bootstrapped standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 2009 CPS data on 25,584 workers in 202 small CBSAs and 34,999 workers in 21 large CBSAs.

would lead to over 900 parameters in each city group, which are hard to identify with our data. We therefore use 2-digit industry classification (52 classes), and 2-digit occupation codes (22 categories), which are both assigned by the NBER. Age enters as a second order polynomial and race in 4 groups (White, Black, Asian, Other).

Table 1 shows the results of the two approaches. The first 3 rows report the raw sample quantiles of the skill, log(utility), distributions as in the right panel of Figure 1. Both methods predict these quantiles well. With the method of Firpo, Fortin, Lemieux (2009), only about 26% of the predicted difference in the 10th percentile can be explained by the composition of observed characteristics. However, 46% of the predicted difference in the 90th percentile can be explained by observed characteristics. Most of this

Table 2: Quantifying the Production Technology.

Observed model outcomes:							
city j	w_{1j}	w_{2j}	w_{3j}	m_{1j}	m_{2j}	m_{3j}	C_j
1	416	844	1923	730,509	1,953,303	730,509	21
2	354	717	1634	30,900	105,516	30,900	204
Implied production technology for different values of γ:							
γ	λ	A_1	A_2	y_1	y_2	y_3	
0.655	1.0407	190,228	59,107	0.2329	1	1.0762	
0.8	1.0193	19,118	9,065	0.3189	1	1.4733	
0.9	1.0086	3,992	2,534	0.3964	1	1.8317	

Notes: w_{ij}: weekly median wage (in US\$); m_{ij}: number of workers of skill type i in cities of type j (in units); C_j: number of cities of type j (in units); γ: exogenously chosen technology parameter; $\lambda, A_j y_i$: estimated technology parameters.

is explained by education (36%) and occupation (17%). The strong explanatory power of observables for the top tail and the relatively low explanatory power for the bottom tail reiterates the findings in the previous sections. The method of Chernozhukov, Fernández-Val, Melly leads to very similar results. 57% of the 90th percentile is explained by the composition of observables and 28% of the 10thpercentile.

The novel finding here is that there is an asymmetry between the low and the high skilled. For the low-skilled, very little of the difference between big and small cities can be explained by observables, whereas for the high-skilled about half can be explained by observables.

8 Quantifying the Production Technology

In this Section, we use the equilibrium properties of the theoretical model to quantify the features of the technology. The model allows us to obtain quantitative information on the technology, in particular the differences in TFP across big and small cities, the productivity differences across differentially skilled workers as well as the magnitude of the extreme-skill complementarity.

Using the observed skill distributions, we perform a very simple quantitative exercise to get an idea of the magnitude of these underlying parameter values of the production technology. To that end, we can actually solve the system of equations explicitly (see the appendix for the derivation) to obtain a system of 5 equations in 5 unknowns $\lambda, A_1, A_2, y_1, y_3$ where y_2 is normalized to 1 and γ is exogenously chosen.[24]

[24] In a CES technology, there is an indeterminacy between the A_j's and the y_i's. Here, since the skill $i = 2$ is CES with the composite of skills 1 and 3, it is proportional to A_j (see the second equation), we can normalize it to 1.

To adapt the data to our model with three skill types and two cities, we partition the distribution of our wage-based measure of skills into three types $i = 1, 2, 3$ corresponding to the 20-60-20 percentiles, and construct two city types $j = 1, 2$; type 1 with population larger than 2.5 million and city type 2 with population less than 1 million. We then use observable median[25] wages w_{ij} within each city type and skill group, the actual number of agents of each type in each city m_{ij},[26] and the number of cities in the sample C_j. Housing prices p_j satisfy equation (4).[27] We use the parameter values $\alpha = 0.24$ (the share of expenditure on housing) and estimate the production technology parameters for different values of γ. The results are reported in Table 2. As will become clear next, we tend to find more reasonable estimates for values of γ closer to 1.

For $\gamma = 0.655$ which corresponds to an elasticity of substitution between low- and high-skill workers of 2.9 (Acemoglu and Autor, 2012), TFP is 3 times higher in the large city than in the small city, while the top skill group is 4.6 times as productive as the bottom skill group, but merely 8% more productive than the middle skill types. The parameter λ that measures the extent of the complementarities/substitutabilities is larger than 1, confirming that there are complementarities between the extreme skill types. The magnitude of λ is 1.041. Instead, for higher γ there is less curvature on the amount of labor in output produced, and as a result TFP differences are much closer. For $\gamma = 0.8$, TFP is double, and the high-skilled are still 4.6 times more productive than the low-skilled (this is also the case for $\gamma = 0.9$) but now they are 47% more productive than the middle-skilled. For $\gamma = 0.9$, TFP is 57% higher in the large cities, and the high-skilled are 80% more productive relative to the middle-skilled types. The technology seems to more reasonably capture the TFP differences and productivity y_3 for high γ, corresponding to high elasticities of substitution. This is consistent with the fact that we do not condition on age (or other observables for that matter), where the age elasticity within skill group is typically large (of the order of 5).

In summary, quantitatively we find that the productivity difference between the top and bottom skilled workers is substantial and in the order of at least four times bigger. For all specifications the degree of extreme-skill complementarity is positive, ranging from 0.8% to 4% depending on the returns to scale. In all specifications, the differences in TFP across large and small cities is big. Even the lowest estimate with limited decreasing returns, TFP is about sixty percent higher in large cities.

[25] We use median wages because the tails are truncated lognormals, and thus heavily skewed. We adjust the observed wages such that the relative wages of different skill types are constant across cities as the theory predicts.

[26] We impose symmetry on the observed numbers for skill types 1 and 3.

[27] The observed prices for the two city groups are very similar to that implied by relative wages.

9 Conclusion

We propose a tractable theory of spatially dispersed production with perfectly mobile heterogeneous inputs, i.e. skilled labor. Differences in TFP lead to differences in demand for skills across cities. In general equilibrium, wages and housing 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. We consider two alternative hypotheses concerning differential complementarities/substitutabilities between skills and derive the implication for the equilibrium skill distribution across cities. First, when there are complementarities between extreme skills, the firm size distribution in larger cities has thicker tails. Instead, when there are complementarities between the top skills, there is first order stochastic dominance of the skill distribution in large cities.

We also find robust empirical evidence from US data for thick tails in the skill distribution. Adjusting wages for housing prices by means of a hedonic price index, we find that average skills are constant, but the standard deviation increases with city size. Big cities have big real inequality. Given the theory, this provides empirical support for the extreme-skill complementarity hypothesis: the productivity of the high-skilled is enhanced most by the providers of low-skilled services.

These findings contribute to our understanding of the urban wage premium. Not only do we establish robustly that higher wages are not due to higher average skill, but we also find that there is an urban inequality "premium". In the presence of extreme skill complementarities, this indicates that extreme skills multiply total factor productivity differences.

Finally, our method and results can provide new insights into the role of complementarities in production. At the economy wide level, we know remarkably little about the skill composition across firms of different sizes, for example, and even less about the pattern of complementarities between differentially skilled workers within firms. Understanding the patterns of complementarities is not only important for the efficient allocation of resources. As we have demonstrated in this paper, they are also key for the equitable distribution of the output of production.

Appendix A: Theory

Preliminaries

The full system of equations that pins down the equilibrium allocation can be written as, for $j \in \{1, 2\}$:

$$\begin{cases} \lambda A_j \left[m_{3j}^{\gamma} y_3 + m_{1j}^{\gamma} y_1\right]^{\lambda-1} \gamma m_{1j}^{\gamma-1} y_1 - w_{1j} = 0 \\ \gamma A_j m_{2j}^{\gamma-1} y_2 - w_{2j} = 0 \\ \lambda A_j \left[m_{3j}^{\gamma} y_3 + m_{1j}^{\gamma} y_1\right]^{\lambda-1} \gamma m_{3j}^{\gamma-1} y_3 - w_{3j} = 0 \\ \left(\frac{w_{i1}}{w_{i2}}\right) = \left(\frac{p_1}{p_2}\right)^{\alpha}, \ \forall i \in \{1, 2, 3\} \\ C_1 m_{i1} + C_2 m_{i2} = M_i, \, \forall i \in \{1, 2, 3\} \\ h_{ij} = \frac{\alpha w_{ij}}{p_j}, \, \forall i \in \{1, 2, 3\} \\ \sum_{i=1}^{3} h_{ij} m_{ij} = H \end{cases} \tag{18}$$

Since $w_{i2} = \left(\frac{p_2}{p_1}\right)^{\alpha} w_{i1}$, we can equate the first-order conditions to obtain:

$$\begin{cases} A_1 \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda-1} m_{11}^{\gamma-1} = \left(\frac{p_1}{p_2}\right)^{\alpha} A_2 \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda-1} m_{12}^{\gamma-1} \\ A_1 m_{21}^{\gamma-1} = \left(\frac{p_1}{p_2}\right)^{\alpha} A_2 m_{22}^{\gamma-1} \\ A_1 \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda-1} m_{31}^{\gamma-1} = \left(\frac{p_1}{p_2}\right)^{\alpha} A_2 \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda-1} m_{32}^{\gamma-1} \\ C_1 m_{i1} + C_2 m_{i2} = M_i, \, \forall i \in \{1, 2, 3\} \\ m_{2j}^{\gamma} y_2 + \lambda \left[m_{3j}^{\gamma} y_3 + m_{1j}^{\gamma} y_1\right]^{\lambda} = H \frac{p_j}{\alpha \gamma A_j} \end{cases} \tag{19}$$

From the first and the third equation, we obtain:

$$\frac{A_1 \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda-1} m_{11}^{\gamma-1}}{A_1 \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda-1} m_{31}^{\gamma-1}} = \frac{\left(\frac{p_1}{p_2}\right)^{\alpha} A_2 \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda-1} m_{12}^{\gamma-1}}{\left(\frac{p_1}{p_2}\right)^{\alpha} A_2 \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda-1} m_{32}^{\gamma-1}} \tag{20}$$

After rearranging and using market clearing ($C_1 m_{i1} + C_2 m_{i2} = M_i$) we can write this as:

$$m_{11} = \frac{M_1}{M_3} m_{31}, \tag{21}$$

and we can write the first equation in (19) as:

$$\frac{\left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda-1} m_{11}^{\gamma-1}}{m_{21}^{\gamma-1}} = \frac{\left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda-1} m_{12}^{\gamma-1}}{m_{22}^{\gamma-1}}. \tag{22}$$

Now given the symmetry assumption $m_{11} = m_{31}$ and $m_{12} = m_{32}$, this then implies:

$$m_{21} = \left(\frac{m_{11}}{m_{12}}\right)^{\frac{\lambda\gamma-1}{\gamma-1}} m_{22}, \tag{23}$$

and we substitute it back in (19) and rearrange to get:

$$\left(\frac{m_{11}}{m_{12}}\right)^{\lambda\gamma-1} = \left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1} \tag{24}$$

Using the fact that $m_{12} = \frac{M_1}{C_2} - \frac{C_1}{C_2}m_{11}$, we have:

$$m_{11} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}\frac{M_1}{C_2}}{1+\frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}} \quad \text{and} \quad m_{12} = \frac{\frac{M_1}{C_2}}{1+\frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}} \tag{25}$$

and likewise for the other expressions for m_{2j} and m_{3j}:

$$m_{21} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma-1}}\frac{M_2}{C_2}}{1+\frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma-1}}} \quad \text{and} \quad m_{22} = \frac{\frac{M_2}{C_2}}{1+\frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma-1}}} \tag{26}$$

$$m_{31} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}\frac{M_3}{C_2}}{1+\frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}} \quad \text{and} \quad m_{32} = \frac{\frac{M_3}{C_2}}{1+\frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}} \tag{27}$$

Substituting the m_{ij}'s in the last equation in (19), and rearranging, we get:

$$\left\{\begin{array}{l} \left[\left(\frac{p_2}{p_1}\right)^{\frac{1-\lambda\gamma(1-\alpha)}{1-\lambda\gamma}}\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\lambda\gamma}}-1\right]\lambda\left(\frac{\frac{M_1}{C_2}}{1+\frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}}\right)^{\lambda\gamma}[y_3+y_1]^{\lambda} \\ \quad +\left[\left(\frac{p_2}{p_1}\right)^{\frac{1-\gamma(1-\alpha)}{1-\gamma}}\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma}}-1\right]\left(\frac{\frac{M_2}{C_2}}{\left\{1+\frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\gamma-1}}\right\}}\right)^{\gamma} y_2 \end{array}\right\}\left(\frac{p_1}{p_2}\frac{A_2}{A_1}\right) = 0 \tag{$\bigstar$}$$

where we have used the fact that $\gamma < 1$, $\lambda > 1$, and $\lambda\gamma < 1$. We can now establish Lemma 1.

Lemma 1 *Let $A_1 > A_2$ and $\lambda\gamma < 1, \gamma < 1$. Then housing prices in the more productive city are larger, $p_1 > p_2$.*

Proof. In order to satisfy the equality ($\bigstar$), the only terms that can be negative are the ones in between squared brackets. Since $\frac{A_1}{A_2} > 1$ and $\min\left\{\frac{1}{1-\lambda\gamma}, \frac{1}{1-\gamma}\right\} > 1$, the only way one of these terms is negative is if

$$\min\left\{\left(\frac{p_2}{p_1}\right)^{\frac{1-\lambda\gamma(1-\alpha)}{1-\lambda\gamma}}\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\lambda\gamma}}, \left(\frac{p_2}{p_1}\right)^{\frac{1-\gamma(1-\alpha)}{1-\gamma}}\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\gamma}}\right\} < 1.$$

However, since $\frac{1-\lambda\gamma(1-\alpha)}{1-\lambda\gamma}$ and $\frac{1-\gamma(1-\alpha)}{1-\gamma}$ are positive, this is only possible if $\frac{p_2}{p_1} < 1 \Rightarrow p_2 < p_1$. ∎

Proof of Theorem 1

Theorem 1 *City Size and TFP. Let $A_1 > A_2$ and $\lambda\gamma < 1, \gamma < 1$. Then the more productive city is larger, $S_1 > S_2$.*

Proof. Based on Lemma 1, we know that $p_1 > p_2$. Since $\lambda > 1$ and $\lambda\gamma < 1$, we have that $\frac{1}{1-\lambda\gamma} > \frac{1}{1-\gamma}1$ and $\frac{1-\lambda\gamma(1-\alpha)}{1-\lambda\gamma} > \frac{1-\gamma(1-\alpha)}{1-\gamma} > 1$. Since we know that $A_1 > A_2$, we have that the first term in squared brackets in ($\bigstar$) is positive if

$$\frac{p_2}{p_1} > \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\lambda\gamma(1-\alpha)}} \tag{28}$$

while the second term in squared brackets is positive if:

$$\frac{p_2}{p_1} > \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\gamma(1-\alpha)}}. \tag{29}$$

Since $\frac{A_2}{A_1} < 1$, we have that:

$$\left(\frac{p_2}{p_1}\right)^{\alpha}\frac{A_1}{A_2} \in \left(\left(\frac{A_1}{A_2}\right)^{\frac{(1-\lambda\gamma)(1-\alpha)}{1-\lambda\gamma(1-\alpha)}}, \left(\frac{A_1}{A_2}\right)^{\frac{(1-\gamma)(1-\alpha)}{1-\gamma(1-\alpha)}}\right) > 1. \tag{30}$$

From the expressions for m_{ij}:

$$m_{11} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}\frac{M_1}{C_2}}{1+\frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}} \tag{31}$$

$$m_{12} = \frac{\frac{M_1}{C_2}}{1+\frac{C_1}{C_2}\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda\gamma-1}}}, \tag{32}$$

and $\left(\frac{p_2}{p_1}\right)^{\alpha}\frac{A_1}{A_2} > 1$, we have that $m_{11} > m_{12}$, and likewise $m_{21} > m_{22}$ and $m_{31} > m_{32}$. Finally, since:

$$S_j = m_{1j} + m_{2j} + m_{3j} \tag{33}$$

it immediately follows that $S_1 > S_2$. ∎

Proof of Theorem 2

Theorem 2 *thick tails. Given $A_1 > A_2$, $\lambda > 1$ and $\lambda\gamma < 1$, the skill distribution in the larger city has thicker tails.*

Proof. Consider the distributions, denoted by $pdf_{ij} = \frac{m_{ij}}{S_j}$, where we denote by $Z = \left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1} < 1$. Then we can write

$$
\begin{aligned}
pdf_{11} &= \frac{\frac{Z^{\frac{1}{\lambda\gamma-1}}\frac{M_1}{C_2}}{1+\frac{C_1}{C_2}Z^{\frac{1}{\lambda\gamma-1}}}}{\frac{Z^{\frac{1}{\lambda\gamma-1}}\frac{M_1+M_3}{C_2}}{1+\frac{C_1}{C_2}Z^{\frac{1}{\lambda\gamma-1}}}+\frac{Z^{\frac{1}{\gamma-1}}\frac{M_2}{C_2}}{1+\frac{C_1}{C_2}Z^{\frac{1}{\gamma-1}}}} \qquad (34)\\
pdf_{12} &= \frac{\frac{\frac{M_1}{C_2}}{1+\frac{C_1}{C_2}Z^{\frac{1}{\lambda\gamma-1}}}}{\frac{\frac{M_1+M_3}{C_2}}{1+\frac{C_1}{C_2}Z^{\frac{1}{\lambda\gamma-1}}}+\frac{\frac{M_2}{C_2}}{1+\frac{C_1}{C_2}Z^{\frac{1}{\gamma-1}}}} \qquad (35)
\end{aligned}
$$

Then:

$$
\frac{pdf_{11}}{pdf_{12}} = \frac{Z^{\frac{1}{\lambda\gamma-1}}\left\{(M_1+M_3)\times\left[C_2+C_1Z^{\frac{1}{\gamma-1}}\right]+M_2\times\left[C_2+C_1Z^{\frac{1}{\lambda\gamma-1}}\right]\right\}}{Z^{\frac{1}{\lambda\gamma-1}}(M_1+M_3)\times\left[C_2+C_1Z^{\frac{1}{\gamma-1}}\right]+\left[C_2+C_1Z^{\frac{1}{\lambda\gamma-1}}\right]\times Z^{\frac{1}{\gamma-1}}M_2} > 1 \qquad (36)
$$

Recall that $Z<1$, and therefore $Z^{\frac{1}{\lambda\gamma-1}}=\left(\frac{1}{Z}\right)^{\frac{1}{1-\lambda\gamma}}$, then since $\frac{1}{Z}>1$, the larger the exponent, the larger is $\left(\frac{1}{Z}\right)^{\frac{1}{1-\lambda\gamma}}$. Since $\lambda>1$ and $\lambda\gamma<1$, $\frac{1}{1-\lambda\gamma}>\frac{1}{1-\lambda}$, it follows that $pdf_{11}>pdf_{12}$.

Similarly we can show that $pdf_{31}>pdf_{32}$:

$$
\frac{pdf_{31}}{pdf_{32}} = \frac{Z^{\frac{1}{\lambda\gamma-1}}\left\{(M_1+M_3)\times\left[C_2+C_1Z^{\frac{1}{\gamma-1}}\right]+M_2\times\left[C_2+C_1Z^{\frac{1}{\lambda\gamma-1}}\right]\right\}}{Z^{\frac{1}{\lambda\gamma-1}}(M_1+M_3)\times\left[C_2+C_1Z^{\frac{1}{\gamma-1}}\right]+\left[C_2+C_1Z^{\frac{1}{\lambda\gamma-1}}\right]\times Z^{\frac{1}{\gamma-1}}M_2} > 1 \qquad (37)
$$

Finally:

$$
\begin{aligned}
pdf_{21} &= \frac{\left[C_2+C_1Z^{\frac{1}{\lambda\gamma-1}}\right]Z^{\frac{1}{\gamma-1}}M_2}{Z^{\frac{1}{\lambda\gamma-1}}(M_1+M_3)\times\left[C_2+C_1Z^{\frac{1}{\lambda-1}}\right]+Z^{\frac{1}{\gamma-1}}M_2\times\left[C_2+C_1Z^{\frac{1}{\lambda\gamma-1}}\right]}\\
pdf_{22} &= \frac{\left[C_2+C_1Z^{\frac{1}{\lambda\gamma-1}}\right]M_2}{(M_1+M_3)\left[C_2+C_1Z^{\frac{1}{\gamma-1}}\right]+M_2\left[C_2+C_1Z^{\frac{1}{\lambda\gamma-1}}\right]}
\end{aligned}
$$

Then:

$$
\frac{pdf_{21}}{pdf_{22}} = \frac{Z^{\frac{1}{\gamma-1}}(M_1+M_3)\left[C_2+C_1Z^{\frac{1}{\lambda-1}}\right]+M_2\left[C_2+C_1Z^{\frac{1}{\lambda\gamma-1}}\right]}{Z^{\frac{1}{\lambda\gamma-1}}(M_1+M_3)\times\left[C_2+C_1Z^{\frac{1}{\lambda-1}}\right]+Z^{\frac{1}{\lambda-1}}M_2\times\left[C_2+C_1Z^{\frac{1}{\lambda\gamma-1}}\right]} \qquad (38)
$$

Again, since $Z^{\frac{1}{\gamma-1}}<Z^{\frac{1}{\lambda\gamma-1}}$, now we have that $pdf_{21}<pdf_{22}$. ■

Proof of Theorem 3

Theorem 3 *Top-Skill Complementarity. Given $A_1>A_2$, $\lambda>1$ and $\lambda\gamma<1$, the skill distribution in the larger city first order stochastically dominates.*

Proof. The proof follows closely the logic of Theorem 2.

$$
\begin{aligned}
pdf_{11} &= \frac{\left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right] Z^{\frac{1}{\gamma-1}} M_1}{Z^{\frac{1}{\lambda\gamma-1}}\left(M_2+M_3\right) \times \left[C_2 + C_1 Z^{\frac{1}{\lambda-1}}\right] + Z^{\frac{1}{\gamma-1}} M_1 \times \left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right]} \\
pdf_{12} &= \frac{\left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right] M_1}{\left(M_2+M_3\right)\left[C_2 + C_1 Z^{\frac{1}{\gamma-1}}\right] + M_1\left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right]}
\end{aligned}
$$

Then:

$$
\frac{pdf_{11}}{pdf_{12}} = \frac{Z^{\frac{1}{\gamma-1}}\left(M_2+M_3\right)\left[C_2 + C_1 Z^{\frac{1}{\lambda-1}}\right] + M_1\left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right]}{Z^{\frac{1}{\lambda\gamma-1}}\left(M_2+M_3\right) \times \left[C_2 + C_1 Z^{\frac{1}{\lambda-1}}\right] + Z^{\frac{1}{\lambda-1}} M_1 \times \left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right]} < 1, \tag{39}
$$

since $Z^{\frac{1}{\gamma-1}} < Z^{\frac{1}{\lambda\gamma-1}}$.

$$
\begin{aligned}
pdf_{21} &= \frac{\left[C_2 + C_1 Z^{\frac{1}{\gamma-1}}\right] Z^{\frac{1}{\lambda\gamma d-1}} M_2}{Z^{\frac{1}{\lambda\gamma-1}}\left(M_2+M_3\right) \times \left[C_2 + C_1 Z^{\frac{1}{\gamma-1}}\right] + \left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right] \times Z^{\frac{1}{\gamma-1}} M_1} \\
pdf_{22} &= \frac{\left[C_2 + C_1 Z^{\frac{1}{\gamma-1}}\right] M_2}{\left(M_2+M_3\right) \times \left[C_2 + C_1 Z^{\frac{1}{\gamma-1}}\right] + M_1 \times \left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right]}
\end{aligned}
$$

Then:

$$
\frac{pdf_{21}}{pdf_{22}} = \frac{Z^{\frac{1}{\lambda\gamma-1}}\left(M_2+M_3\right) \times \left[C_2 + C_1 Z^{\frac{1}{\gamma-1}}\right] + M_1 \times \left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right]}{Z^{\frac{1}{\lambda\gamma-1}}\left(M_2+M_3\right) \times \left[C_2 + C_1 Z^{\frac{1}{\gamma-1}}\right] + \left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right] \times Z^{\frac{1}{\gamma-1}} M_1} > 1 \tag{40}
$$

Finally, analogously to $\frac{pdf_{21}}{pdf_{22}}$, we can derive $\frac{pdf_{31}}{pdf_{32}}$:

$$
\frac{pdf_{31}}{pdf_{32}} = \frac{Z^{\frac{1}{\lambda\gamma-1}}\left(M_1+M_3\right) \times \left[C_2 + C_1 Z^{\frac{1}{\gamma-1}}\right] + M_2 \times \left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right]}{Z^{\frac{1}{\lambda\gamma-1}}\left(M_1+M_3\right) \times \left[C_2 + C_1 Z^{\frac{1}{\gamma-1}}\right] + \left[C_2 + C_1 Z^{\frac{1}{\lambda\gamma-1}}\right] \times Z^{\frac{1}{\gamma-1}} M_2} > 1 \tag{41}
$$

■

Quantifying the Production Technology: Derivation of the System

The equilibrium allocation can explicitly be represented by the following system of 5 equations in 5 unknowns $\lambda, A_1, A_2, y_1, y_3$ where y_2 is normalized to 1:

$$\lambda = \frac{1}{\gamma}\left[1 + (\gamma - 1)\frac{\log\left(\frac{C_2 m_{21}}{M_2 - C_1 m_{21}}\right)}{\log\left(\frac{C_2 m_{11}}{M_1 - C_1 m_{11}}\right)}\right], \quad A_1 = \frac{w_{21}}{\gamma y_2 m_{21}^{\gamma-1}}, \quad A_2 = A_1\left(\frac{p_2}{p_1}\right)^{\alpha}\left(\frac{C_2 m_{21}}{M_2 - C_1 m_{21}}\right)^{\gamma-1}, \tag{42}$$

$$y_1 = \left(\frac{w_{11}}{\lambda\gamma A_1\left[m_{11} + m_{31}\frac{w_{31}}{w_{11}}\right]^{\lambda-1} m_{11}^{\lambda(\gamma-1)}}\right)^{\frac{1}{\lambda}}, \quad y_3 = \left(\frac{w_{31}}{\lambda\gamma A_1\left[m_{31} + m_{11}\frac{w_{11}}{w_{31}}\right]^{\lambda-1} m_{31}^{\lambda(\gamma-1)}}\right)^{\frac{1}{\lambda}}$$

From equation (26), we can solve for Z:

$$Z = \left(\frac{C_2 m_{21}}{M_2 - C_1 m_{21}}\right)^{\gamma-1}. \tag{43}$$

Now solve the expression for m_{11} in terms of λ, and substitute for this expression for Z:

$$\lambda = \frac{1}{\gamma}\left[1 + (\gamma - 1)\frac{\log\left(\frac{C_2 m_{21}}{M_2 - C_1 m_{21}}\right)}{\log\left(\frac{C_2 m_{11}}{M_1 - C_1 m_{11}}\right)}\right]. \tag{44}$$

Therefore, the solution for λ, Z is simply the explicit solution to the equations (43) and (44).

Now from the second FOC (equation (10)) we obtain:

$$y_2 = \frac{w_{21}}{\gamma A_1 m_{21}^{\gamma-1}}, \tag{45}$$

while from the first (equation (9)) and the third FOC (equation (11)), we can solve explicitly for y_1 and y_3:

$$\begin{aligned} y_1 &= \frac{w_{11}}{w_{31}}\frac{m_{31}^{\gamma-1}}{m_{11}^{\gamma-1}}\left(\frac{w_{31}}{\lambda\gamma A_1\left[m_{31} + m_{11}\frac{w_{11}}{w_{31}}\right]^{\lambda-1} m_{31}^{\lambda(\gamma-1)}}\right)^{\frac{1}{\lambda}} \\ &= \left(\frac{w_{11}}{\lambda\gamma A_1\left[m_{11} + m_{31}\frac{w_{31}}{w_{11}}\right]^{\lambda-1} m_{11}^{\lambda(\gamma-1)}}\right)^{\frac{1}{\lambda}} \end{aligned} \tag{46}$$

$$y_3 = \left(\frac{w_{31}}{\lambda\gamma A_1\left[m_{31} + m_{11}\frac{w_{11}}{w_{31}}\right]^{\lambda-1} m_{31}^{\lambda(\gamma-1)}}\right)^{\frac{1}{\lambda}}. \tag{47}$$

Finally, we obtain the expression for A_1 from the second FOC (10) and the expression for A_2 from substituting for $Z = \left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}$ in (43):

$$A_1 = \frac{w_{21}}{\gamma y_2 m_{21}^{\gamma-1}} \tag{48}$$

$$A_2 = A_1 \left(\frac{p_2}{p_1}\right)^{\alpha} \left(\frac{C_2 m_{21}}{M_2 - C_1 m_{21}}\right)^{\gamma-1}. \tag{49}$$

Normalizing $y_2 = 1$, equations (44), (48), (49), (46), and (47) constitute the system (42).

Table 3: 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
...		
245	Farmington, NM	124,131
246	Bowling Green, KY	120,595
247	Harrisonburg, VA	120,271
248	Lawrence, KS	116,383
249	Victoria, TX	115,396
250	Anniston-Oxford, AL	114,081
251	Lawton, OK	113,228
252	Kankakee-Bradley, IL	113,215
253	Michigan City-La Porte, IN	111,063
254	Decatur, IL	108,204

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 ones with population above 100,000 and where housing prices are observed.

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.[28] 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).[29] 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.[30] This variable reports earnings per week in the

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

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

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

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?") appearshere.

The NBER version of the CPS identifies the core-based statistical area (CBSA) of the observation. It reports 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. Table 3 shows the 10 largest and 10 smallest included CBSAs.

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 76,821 workers in 254 identified CBSAs 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 2,308 or 3.0% of workers. All estimations use the weights in variable 'earnwt' provided by the NBER.

Our baseline results use CPS wage data because they are generally considered of higher quality than Census data (see e.g. Baum-Snow and Neal, 2009). However, the CPS has two disadvantages: it has relatively low top codes and it does not identify the location of the household *within* cities. We therefore alternatively use wage data from the 2009 American Community Survey (ACS) collected by the U.S. Census Bureau.[31] The data is provided by the Minnesota Population Center in its Integrated Public Use Microdata Series (IPUMS).[32]

We use the variable 'incwage' which measures yearly wage and salary income. We restrict our sample to full-time (between 36 and 60 usual hours per week) and full-year (between 48 and 52 weeks per year) workers. The yearly wage is divided by the number of weeks worked to get weakly wages comparable to the CPS data. ACS are top-coded at the 99.5 percentile of each state. We also drop the lowest 0.1% of wages as a pragmatic way of eliminating likely misreported wages close to zero. Our final ACS wage sample includes 654,043 workers in 293 identified CBSAs.

The ACS 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) [33] to link PUMA areas to CBSAs. The MCDC data matches every urban PUMA code to one or more CBSA codes and reports the fraction of housing units that are matched. We assign a PUMA to a CBSA if this fraction is bigger than 33%. In cases where the PUMA does not fully belong to a CBSA, we assign the PUMA to the CBSA where most of its housing units belong to. Our final sample contains data from 533 metropolitan or micropolitan core based statistical areas (CBSA) out of a total of 940 existing CBSAs. Note that we do *not* use the metropolitan area code provided in the ACS 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. Our final sample contains 273,761 rental units in 533 CBSAs and 1884 PUMA areas.

[31] See http://www.census.gov/acs/www/ for more information on the survey.

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

[33] Available at http://mcdc.missouri.edu/websas/geocorr2k.html.

Local housing and commodity price indices

We use the 2009 American Community Survey (ACS) for our baseline housing price estimates.

The variable 'rent' reports the monthly contract rent for rental units in contemporary dollars. We also use all the reported housing characteristics of the unit: 'rooms' is the number of rooms, 'unitsstr' is the units in structure (in 8 groups), and 'builtyr' is the age of structure (in 13 age groups).

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).

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 302 core-based statistical areas (CBSA) and 23 metropolitan divisions for the 11 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.

Hedonic Regression to calculate housing price index

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

$$u(c, h(z)) = c^{1-\alpha} h^{\alpha}(z) \tag{50}$$

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

$$h(z) = z_1^{\delta} z_2^{1-\delta}. \tag{51}$$

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

$$U_i = \alpha^{\alpha} (1-\alpha)^{1-\alpha} \left[L q_1^{\delta} q_2^{1-\delta} \right]^{-\alpha} w \tag{52}$$

Table 4: Hedonic regressions for rental units.

	CBSA level		PUMA level	
Number of rooms				
1	-0.2314***	(0.0056)	-0.2238***	(0.0055)
2	-0.1658***	(0.0050)	-0.1863***	(0.0049)
3	-0.1329***	(0.0031)	-0.1386***	(0.0030)
4	0		0	
5	0.0760***	(0.0033)	0.0798***	(0.0031)
6	0.1614***	(0.0041)	0.1592***	(0.0039)
7	0.2405***	(0.0057)	0.2313***	(0.0055)
8	0.2877***	(0.0077)	0.2717***	(0.0074)
9+	0.3341***	(0.0082)	0.3049***	(0.0079)
Age of structure				
1939 or earlier	-0.3068***	(0.0053)	-0.2700***	(0.0053)
1940-1949	-0.3603***	(0.0062)	-0.3219***	(0.0061)
1950-1959	-0.3167***	(0.0055)	-0.2970***	(0.0054)
1960-1969	-0.2887***	(0.0053)	-0.2793***	(0.0052)
1970-1979	-0.2553***	(0.0050)	-0.2542***	(0.0049)
1980-1989	-0.1758***	(0.0052)	-0.1838***	(0.0050)
1990-1999	-0.0780***	(0.0054)	-0.0838***	(0.0052)
2000-2004	0		0	
2005	0.0122	(0.0097)	0.0223**	(0.0094)
2006	0.0421***	(0.0099)	0.0537***	(0.0095)
2007	0.0548***	(0.0104)	0.0621***	(0.0100)
2008	0.1029***	(0.0135)	0.1139***	(0.0130)
2009	0.0343	(0.0444)	0.0347	(0.0427)
Units in structure				
1-family house detached	0		0	
1-family house attached	-0.0635***	(0.0050)	-0.0677***	(0.0049)
2-family building	-0.1257***	(0.0045)	-0.1289***	(0.0044)
3-4 family building	-0.1314***	(0.0042)	-0.1434***	(0.0041)
5-9 family building	-0.1239***	(0.0042)	-0.1532***	(0.0041)
10-19 family building	-0.0786***	(0.0043)	-0.1171***	(0.0042)
20-49 family building	-0.1023***	(0.0048)	-0.1354***	(0.0047)
50+ family building	-0.0929***	(0.0045)	-0.1413***	(0.0045)
Constant	6.5728***	(0.0481)	6.0277***	(0.0523)
CBSA Fixed Effects	yes		no	
PUMA Fixed Effects	no		yes	
N (rental units)	273,761		273,761	
Number of CBSAs	533		533	
Number of PUMA regions			1884	

Notes: * p<0.10, ** p<0.05, *** p<0.01. Reference groups are indicated by "0".

where $L = 1/[\delta^\delta (1-\delta)^{1-\delta}]$. Defining the price index $p = Lq_1^\delta q_2^{1-\delta}$ the indirect utility is

$$U_i = \alpha^\alpha (1-\alpha)^{1-\alpha} \frac{w}{p^\alpha} \tag{53}$$

Table 5: Rank of cities by estimated housing price index.

	City	Population	Rent Index
1	San Jose-Sunnyvale-Santa Clara, CA	1,839,700	1.74
2	San Francisco-Oakland-Fremont, CA	4,317,853	1.64
3	Santa Barbara-Santa Maria-Goleta, CA	407,057	1.62
4	Oxnard-Thousand Oaks-Ventura, CA	802,983	1.62
5	Honolulu, HI	907,574	1.61
6	Los Angeles-Long Beach-Santa Ana, CA	12,874,797	1.55
7	San Diego-Carlsbad-San Marcos, CA	3,053,793	1.51
8	Washington-Arlington-Alexandria, DC-VA-MD-WV	5,476,241	1.46
9	Napa, CA	134,650	1.43
10	Santa Cruz-Watsonville, CA	256,218	1.43
11	New York-Northern New Jersey-Long Island, NY-NJ-PA	19,069,796	1.41
...			
245	McAllen-Edinburg-Mission, TX	741,152	.50
246	Lawton, OK	113,228	.50
247	Lake Charles, LA	194,138	.49
248	Huntington-Ashland, WV-KY-OH	285,624	.49
249	Monroe, LA	174,086	.47
250	Johnstown, PA	143,998	.47
251	Brownsville-Harlingen, TX	396,371	.47
252	Decatur, AL	151,399	.46
253	Joplin, MO	174,300	.45
254	Anniston-Oxford, AL	114,081	.36

Notes: Housing price indices based on hedonic regressions using the 2009 American Community Survey.

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

$$e(q_1, q_2, h) = Lq_1^{\delta} q_2^{1-\delta} h = ph = pz_1^{\delta} z_2^{1-\delta}. \tag{54}$$

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

$$\log(e_{jn}) = \log(p_j) + \delta \log(z_{1jn}) + u_{jn} \tag{55}$$

where e_{jn} is the observed rental price of housing unit n and $\log(p_j)$ is city-specific common component of housing prices. 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.

Table 4 shows the results of the hedonic regressions for rental units using data from the 2009 American Community Survey (ACS). Column (1) shows the results with 533 fixed effects for cities (CBSA) and column (2) with 1844 fixed effects for neighbourhoods (PUMA areas). We use all relevant housing characteristics in the

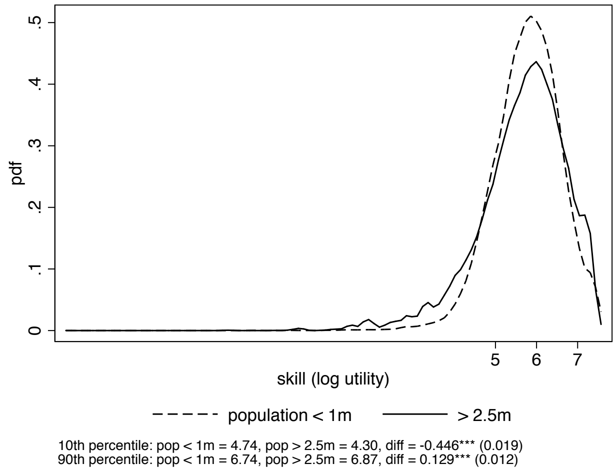

Figure 14: Expenditure Share using Stone-Geary preferences, $\alpha = 0.22$, $\underline{h} = 250$.

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 lowest prices for 3-4 family buildings.

We standardize the housing price index such that the weighted (by housing units) average equals 1. Table 5 shows the resulting housing price indices for the highest and lowest priced cities in our sample. The highest priced city is San Jose, CA with rental prices 74% above average urban prices; the lowest priced city with more than 100,000 inhabitants is Anniston, AL with prices 64% below average.

Expenditure Data for Estimation of Non-Linear Engel Curve

We use Public Use Microdata (PUMD) from the Consumer Expenditure Survey (CEX) which is provided by the U.S. Bureau of Labor Statistics.[34] Our sample consists of interview data for 4581 households in the first quarter of 2011 in the data file "fmli111x". The PUMD reports the exact city (MSA) for 2021 households living in one of the 21 largest MSAs. We merge these data with our hedonic price indices from the corresponding CBSAs. The housing expenditure shares, s_i, are measured as "expenditures on shelter this quarter" (variable SHELTCQ) divided by "total expenditures this quarter" (TOTEXPCQ). Weekly wages, w_i are measured as "income before taxes in past 12 months" (FINCBTAX) divided by 52. Housing prices, p_j, are measured as the hedonic price index using the 2009 American Community Survey. We restrict our sample to weekly wage incomes above $175 because very low wages become extremely large inverse values which would almost fully determine the regression. The limit of $175 is also the lower bound in our CPS wage data (see the data section on wages) and guarantees that the indirect utility is defined for all observations. The regression with 1569 observations results in: $\widehat{\alpha} = 0.224$

[34] See http://www.bls.gov/cex/ for details on the CEX and its public use individual data version. The CEX allows us to calculate expenditure share as a ratio of total expenditures. The ACS does not provide total expenditures of individuals or households. Housing shares as a ratio of reported income are extremely noisy with housing expenditure shares above 100% for a large fraction of low income households. We did not get any reasonable Stone-Geary parameters based on ACS data.

(s.e.$=$ 0.005), $\widehat{\beta} = 21.5$ (3.05), $\widehat{\underline{h}} = 27.7$ (3.79 using the delta method). As a robustness check, we also use values for $\underline{h}$ different from the estimated value. As an extreme example, for $\underline{h} = 250$ we find that even though the shape of the distribution changes dramatically – especially for the low income workers – the thick tails are still prominently present, as illustrated in Figure 14.

Appendix C: Preferences for Services and Home Production

In this appendix we explore the micro foundations for the complementarities between extreme skills. We consider an explanation based on preferences for low-skilled services in combination with home production. Households have preferences for low-skilled services. Those services can be produced with home production and they can be traded on the market.

Citizens have preferences over the quantity of a consumption good c, services s and the amount of housing h represented by $u(h,s,c) = h^{\alpha}s^{\beta}c^{1-\alpha-\beta}$ where α, β, and $(\alpha+\beta) \in [0,1]$. All workers have a unit endowment of time to be divided between home production denoted by t and market production $1-t$. Home productivity is independent of the worker's skills. The amount of services generated in home production is equal to Γt^{δ}, where Γ is a positive parameter, but the agent incurs a quasi fixed cost K, a cost that is incurred only if there is positive production. Services can be traded at a city-specific price $r_j \geq 0$. As before, output in the formal sector generates a wage that is contingent on the worker's skill and produced with the technology $A_j F(\cdot)$ that is either CES or satisfies top- or extreme-skill complementarity.

A citizen i in city j chooses the bundle $\{h,s,c\}$ to maximize utility subject to the budget constraint:

$$\begin{aligned} \max_{\{h,s,c,t\}} u(c,h,s) &= h^{\alpha}s^{\beta}c^{1-\alpha-\beta} \\ \text{s.t. } ph + rs + c &\leq w(1-t) + I \end{aligned} \tag{56}$$

where the quantities h,s,c,t,w are all city- and skill-specific and the prices p,r are city-specific, but we have dropped the subscripts for notational convenience. We denote by I the income generated through services provided. The decision problem to produce services t is given by $\max_{0<t\leq 1}\{r\Gamma t^{\delta} - K + wt\}$, if $t > 0$, and 0 otherwise. Observe that K acts as a cost of entry in the services sector. Solving this problem, and given the cost K, the optimal solution satisfies:

$$I = \max_t \left\{ r\Gamma \left(\frac{r\delta\Gamma}{w}\right)^{\frac{\delta}{1-\delta}} - K, 0 \right\}. \tag{57}$$

Because of the fixed cost, there will be an occupational choice decision whether or not to produce services, and if so, how much (t). With this optimal allocation of time between production of services and of market output, we can pin down the demands for consumption, services, and housing as:

$$h = \frac{\alpha}{p}\left[w(1-t^{\star}) + I\right], \quad s = \frac{\beta}{r}\left[w(1-t^{\star}) + I\right], \quad c = (1-\alpha-\beta)\left[w(1-t^{\star}) + I\right] \tag{58}$$

where $t^{\star} = \max\{\left(\frac{r\delta\Gamma}{w}\right)^{\frac{1}{1-\delta}}, 0\}$. Observe that the time spent in household production depends on the ratio r/w. The higher the wage, the less time she spends producing household services, and the more likely she is a net demander of services. Moreover, because of the cost K, those with high enough wages will choose not to produce services at all. Finally, the market clearing conditions (including in the market for services) pin down equilibrium prices and close the model.

Table 6: Preferences for Services and Home Production.

Assumed production technology:								
γ	λ	A_1	A_2	y_1	y_2	y_3	C_1	C_2
0.8	1.0193	19,118	9,065	0.3189	1	1.4733	21	204
Observed model outcomes:								
city j	w_{1j}	w_{2j}	w_{3j}	pdf_{1j}	pdf_{2j}	pdf_{3j}	p_j	r_j
1	430	1,315	2,330	0.2128	0.5693	0.2178	4473.43	653.88
2	261	774	1,419	0.2016	0.5986	0.1998	763.33	372.82

We cannot solve the model analytically. We have therefore performed various quantitative exercises to get an idea of the properties of the model. In particular, we want to find parameter configurations under which we obtain thick tails. We use the parameters $\alpha = 0.24, \beta = 0.2, \delta = 0.3, \gamma = 0.8, \Gamma = 1, K = 0.2$, and from the data $M_1 = 21,644,289, M_2 = 62,544,627, M_3 = 21,644,289$, $C_1 = 21, C_2 = 204$, and $H = 110,016.5$, which is based on the observed mean number of housing units. Then we use the technology parameters generated by the exercise in section 8. The objective is to obtain distributional properties that are consistent with thick tails.

First, we can robustly reproduce the thick tails results in this model with services whenever there are extreme-skill complementarities in the market technology. This confirms that our approach in the baseline model is robust to the introduction of low-skilled services. At the same time, this may not be all that surprising especially whenever the mechanism of extreme-skill complementarities is strong enough. More challenging is whether thick tails obtain without extreme-skill complementarities. Our *second* finding is that with a CES market technology and without the quasi fixed costs K, the distributions with services are identical across cities. With $K > 0$ and CES, the distributions differ across cities, and results in FOSD of the small city: there are relatively more low-skilled workers in the large city, relatively more middle skilled workers in the small city and the same density of high-skilled workers in both cities; the average skills are lower in the large cities. *Third*, the model generates thick tails if preferences for services are combined with top-skill complementarity. The logic is that the top-skill complementarity generates the fat upper tail, whereas the demand for services generates the fat lower tail.

Table 6 shows this third result in a quantitative exercise. To obtain thick tails, the income share of services β must be large enough. When it drops below 20%, the lower tail disappears.[35] This means that households must spend nearly as much on low-skilled services as on housing. Second, the fixed cost K must also be large enough ($K = 0.2$ is a big barrier), which effectively inflates the price of services. Observe that there is no such entry barrier in the formal sector, where one would expect those to be at least as big. Finally, the top skill complementarity must not be too strong.

Our simulations suggest that demand for services can indeed contribute to the fat lower tail and that it is broadly consistent with extreme-skill complementarity. To generate thick tails without extreme-skill complementarity, however, the required parameter are somewhat extreme. In particular, it requires a strong non-homotheticity in the services technology, and the expenditure share on services must be unrealistically high.

[35] In the CEX, the direct expenditure on low-skilled services is around 5%.

Many low-skilled services are demanded *indirectly* because they are inputs in production: cooks in restaurants, administrative staff in firms, etc. We find it therefore justified to model their role through the production technology. Finally, a prime advantage of our basic model with complementarities in the production technology is that we can solve and study it analytically.

References

[1] Acemoglu, Daron, and David Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings," *Handbook of Labor Economics*, 4(B), 2011, 1043-1171.

[2] Autor, David H., Frank Levy, and Richard J. Murnane. "The Skill Content of Recent Technological Change: An Empirical Investigation", *Quarterly Journal of Economics* 118(4), 2003, 1279-1333.

[3] Autor, David H. and and David Dorn. "The Growth of Low-Skill Service Jobs and the Polarization of the U.S. Labor Market", *American Economic Review* 103(5), 2013, 1553–1597.

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

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

[6] Baum-Snow Nathaniel and Derek Neal, "Mismeasurement of Usual Hours Worked in the Census and ACS" *Economics Letters*, 102(1), 2009, 39-41.

[7] Baum-Snow and Ronni Pavan, "Understanding the City Size Wage Gap," *Review of Economic Studies* 79(1), 2012, 88-127.

[8] Behrens, Kristian, Gilles Duranton and Frederic Robert-Nicoud "Productive cities: Sorting, Selection and Agglomeration", *Journal of Policitcal Economy*, forthcoming.

[9] Blinder, Alan, "Wage Discrimination: Reduced Form and Structural Estimates", *Journal of Human Resources* 8, 1973, 436–455.

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

[11] 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, 77-97.

[12] 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.

[13] Chernozhukov, Victor, Ivan Fernandez-Val, and Blaise Melly, "Inference on Counterfactual Distributions," *Econometrica* 81(6), 2013, 2205–2268.

[14] 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.

[15] Davis, Steven J., R. Jason Faberman, and John Haltiwanger, "The Flow Approach to Labor Markets: New Data Sources and Micro–Macro Links," *Journal of Economic Perspectives* 20(3), 2006, 3–26.

[16] Davis, Morris A., Jonas Fisher, and Toni Whited, "Macroeconomics Implications of Agglomeration," 2013 mimeo.

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

[18] De la Roca, Jorge, and Diego Puga, "Learning by working in big cities", CEMFI mimeo 2012.

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

[20] DiNardo, John, Nicole M. Fortin, and Thomas Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica* 64, 1996, 1001-1044.

[21] Duranton, Gilles and Diego Puga "Micro-foundations of urban agglomeration economies," *Handbook of Regional and Urban Economics*, Volume 4, 2004, 2063-2117.

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

[23] Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux , "Unconditional Quantile Regressions", *Econometrica* 77(3), 2009, 953–973.

[24] Fortin, N.M., Lemieux, T. and S. Firpo, Decomposition Methods. In: O. Ashenfelter and D. Card (eds.), Handbook of Labor Economics, Vol. 4A, Amsterdam: North-Holland, 2011, 1–102.

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

[26] Fujita, Masahisa, *Urban Economic Theory. Land Use and City Size,* Cambridge: Cambridge University Press, 1989.

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

[28] Gautier, P.A., Svarer, M., and Teulings, C.N., "Marriage and the City: Search Frictions and Sorting of Singles (with M. Svarer and C.N. Teulings), 2009, Journal of Urban Economics, 67-2, 206-18

[29] Gautier, Pieter, and C.N. Teulings, "Search and the City," *Regional Science and Urban Economics* 39, 2009, 251-265.

[30] Goos, Maarten, Alan Manning, and Anna Salomons "Job Polarization in Europe," *American Economic Review: Papers & Proceedings* 99(2), 2009, 58-63.

[31] Guerrieri, Veronica, Daniel Hartley, and Erik Hurst, "Endogenous Gentrification and Housing Price Dynamics," University Chicago mimeo 2011.

[32] Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan, "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *Journal of Political Economy* 111(3), 2003, 465-497.

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

[34] Juhn, Chinhui, Kevin M. Murphy, Brooks Pierce, “Wage Inequality and the Rise in Returns to Skill,” *The Journal of Political Economy* 101(3), 1993, 410–442.

[35] Kelso, Alexander S, J., and V.P. Crawford “Job Matching, Coalition Formation, and Gross Substitutes,” *Econometrica*, 50, 1982, 1483–1504.

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

[37] Marshall, Alfred “Principles of Economics”, *Macmillan*, 1890, Cambridge, MA.

[38] Moretti, Enrico. 2013. "Real Wage Inequality." *American Economic Journal: Applied Economics* 5(1): 65-103.

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

[40] Oaxaca, Ronald, “Male-Female Wage Differentials in Urban Labor Markets,” *International Economic Review* 14, 1973, 693-709.

[41] 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.

[42] Rothe, Christoph, “Decomposing the Composition Effect: Understanding the Role of Covariates in Determining Between-Group Differences in Economic Outcomes”, mimeo, Columbia University, 2012.

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

[44] Schmidheiny, Kurt "Income Segregation from Local Income Taxation When Households Differ in Both Preferences and Incomes", *Regional Science and Urban Economics*, 36(2), 2006, 270-299.

[45] Van Nieuwerburgh, Stijn, and Pierre-Olivier Weill, “Why Has House Price Dispersion Gone Up?” *Review of Economic Studies*, 77(4), 2010.

Spatial Sorting

Jan Eeckhout, Roberto Pinheiro, and Kurt Schmidheiny

Online Appendix

Journal of Political Economy

I. General Technology: N skill types

The firm's problem with N skills is given by:

$$\pi\left(m_{1j},...,m_{jN}\right)=A_jF\left(m_{1j},...,m_{jN}\right)-\sum_{i=1}^{N}w_{ij}m_{ij} \tag{A.1}$$

Then, the system becomes:

$$\begin{cases} A_1F_i\left(m_{11},...,m_{N1}\right)=\left(\frac{p_1}{p_2}\right)^{\alpha}A_2F_i\left(m_{12},....,m_{N2}\right), & \forall i\in\{1,...,N\}\\ \sum_{i=1}^{N}F_i\left(m_{11},...,m_{N1}\right)m_{i1}=H\frac{p_1}{\alpha A_1}\\ \sum_{i=1}^{N}F_i\left(m_{12},....,m_{N2}\right)m_{i2}=H\frac{p_2}{\alpha A_2}\\ C_1m_{i1}+C_2m_{i2}=M_i, & \forall i\in\{1,...,N\}\end{cases} \tag{A.2}$$

Now, define $F(\cdot)$ as (assuming without loss that N is even):

$$F(\cdot)=\left(m_{11}^{\gamma}y_1+m_{N1}^{\gamma}y_N\right)^{\lambda_1}+\left(m_{21}^{\gamma}y_2+m_{N-1,1}^{\gamma}y_{N-1}\right)^{\lambda_2}+...+\left(m_{\frac{N}{2}1}^{\gamma}y_{\frac{N}{2}}+m_{\frac{N}{2}+1,1}^{\gamma}y_{\frac{N}{2}+1}\right)^{\lambda_{\frac{N}{2}}} \tag{A.3}$$

Substituting this back into the system, we have:

$$\begin{cases} \frac{A_1\left(m_{i,1}^{\gamma}y_i+m_{N-(i-1),1}^{\gamma}y_{N-(i-1)}\right)^{\lambda_{\min\{i,N-(i-1)\}_i}-1}m_{i,1}^{\gamma-1}}{A_2\left(m_{i,2}^{\gamma}y_i+m_{N-(i-1),2}^{\gamma}y_{N-(i-1)}\right)^{\lambda_{\min\{i,N-(i-1)\}_i}-1}m_{i,2}^{\gamma-1}}=\left(\frac{p_1}{p_2}\right)^{\alpha}, \quad \forall i\in\{1,...,N\} & (1),...,(N)\\ \sum_{i=1}^{N}\lambda_{\min\{i,N-(i-1)\}_i}\left(m_{i,1}^{\gamma}y_i+m_{N-(i-1),1}^{\gamma}y_{N-(i-1)}\right)^{\lambda_{\min\{i,N-(i-1)\}_i}-1}m_{i,1}^{\gamma}y_i=H\frac{p_1}{\alpha\beta\gamma A_1} & (N+1)\\ \sum_{i=1}^{N}\lambda_{\min\{i,N-(i-1)\}_i}\left(m_{i,2}^{\gamma}y_i+m_{N-(i-1),2}^{\gamma}y_{N-(i-1)}\right)^{\lambda_{\min\{i,N-(i-1)\}_i}-1}m_{i,2}^{\gamma}y_i=H\frac{p_2}{\alpha\beta\gamma A_2} & (N+2)\\ C_1m_{i1}+C_2m_{i2}=M_i, \quad \forall i\in\{1,...,N\} & (N+3),...,(2N+2)\end{cases} \tag{A.4}$$

From the first N equations, dividing the expressions for i and $N-(i-1)$, we have:

$$m_{i,1}=\frac{M_i}{M_{N-(i-1)}}m_{N-(i-1),1},\ \text{for } i\in\left\{1,...,\frac{N}{2}\right\} \tag{A.5}$$

Considering a symmetric distribution $\left(\text{so } M_i=M_{N-(i-1)},\text{ for } i\in\left\{1,...,\frac{N}{2}\right\}\right)$, we have that $m_{i,1}=m_{N-(i-1),1}$. Similarly $m_{i2}=m_{N-(i-1),2}=\frac{M_i}{N_2}-\frac{N_1}{N_2}m_{i1}$, for $i\in\left\{1,...,\frac{N}{2}\right\}$.

From equations for i and j for $i\neq j$ from the first N equations, we have:

$$\left[\frac{m_{i,1}}{m_{N-(i-1),1}}\right]^{\gamma-1}=\left[\frac{m_{i,2}}{m_{N-(i-1),2}}\right]^{\gamma-1},\text{ for } i\in\left\{1,...,\frac{N}{2}\right\} \tag{A.6}$$

Using equations $(N+3)$ to $(2N+2)$, we have:

$$m_{i,1}=\frac{M_i}{M_{N-(i-1)}}m_{N-(i-1),1},\text{ for } i\in\left\{1,...,\frac{N}{2}\right\}$$

Considering a symmetric distribution $\left(\text{so } M_i=M_{N-(i-1)},\text{ for } i\in\left\{1,...,\frac{N}{2}\right\}\right)$, we have that $m_{i,1}=m_{N-(i-1),1}$. Similarly $m_{i2}=m_{N-(i-1),2}=\frac{M_i}{C_2}-\frac{C_1}{C_2}m_{i1}$, for $i\in\left\{1,...,\frac{N}{2}\right\}$.

From the first N equations, we have:

$$m_{i,1}=\frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda_{\min\{i,N-(i-1)\}_i}\gamma-1}}M_i}{C_2+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda_{\min\{i,N-(i-1)\}_i}\gamma-1}}C_1} \tag{A.7}$$

Similarly:

$$m_{i2}=\frac{M_i}{C_2+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda_{\min\{i,N-(i-1)\}_i}\gamma-1}}C_1} \tag{A.8}$$

Then, from $(N+1)$, we have:

$$\frac{A_1}{p_1}\sum_{i=1}^{N}\left\{\begin{array}{c}\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{\lambda_{\min\{i,N-(i-1)\}_i}\gamma}{\lambda_{\min\{i,N-(i-1)\}_i}\gamma-1}}\left(\frac{M_i}{C_2+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda_{\min\{i,N-(i-1)\}_i}\gamma-1}}C_1}\right)^{\lambda_{\min\{i,N-(i-1)\}_i}\gamma}\\ \times\lambda_{\min\{i,N-(i-1)\}_i}\left(y_i+y_{N-(i-1)}\right)^{\lambda_{\min\{i,N-(i-1)\}_i}-1}y_i\end{array}\right\}=\frac{H}{\alpha\beta\gamma} \tag{A.9}$$

Similarly, from $(N+2)$, we have:

$$\frac{A_2}{p_2}\sum_{i=1}^{N}\left\{\begin{array}{c}\left(\frac{M_i}{C_2+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda_{\min\{i,N-(i-1)\}_i}\gamma-1}}C_1}\right)^{\lambda_{\min\{i,N-(i-1)\}_i}\gamma}\\ \times\lambda_{\min\{i,N-(i-1)\}_i}\left(y_i+y_{N-(i-1)}\right)^{\lambda_{\min\{i,N-(i-1)\}_i}-1}y_i\end{array}\right\}=\frac{H}{\alpha\beta\gamma} \tag{A.10}$$

Combining these expressions and rearranging, we have:

$$\sum_{i=1}^{N}\left\{\begin{array}{c}\left[\left(\frac{p_2}{p_1}\right)^{\frac{1-(1-\alpha)\lambda_{\min\{i,N-(i-1)\}_i}\gamma}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}}\left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}}-1\right]\times\\ \times\left(\frac{M_i}{C_2+\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\right]^{\frac{1}{\lambda_{\min\{i,N-(i-1)\}_i}\gamma-1}}C_1}\right)^{\lambda_{\min\{i,N-(i-1)\}_i}\gamma}\\ \times\lambda_{\min\{i,N-(i-1)\}_i}\left(y_i+y_{N-(i-1)}\right)^{\lambda_{\min\{i,N-(i-1)\}_i}-1}y_i\end{array}\right\}=0\qquad(\bigstar) \tag{A.11}$$

Lemma A.1: Let $\lambda_i>1,\lambda_i\gamma<1$, for every $i\in\left\{1,2,...,\frac{N}{2}\right\}$ and $\{\lambda_i\}_{i=1}^{\frac{N}{2}}$ be a decreasing sequence. If $A_1>A_2$, then house prices are higher in the city with higher TFP, $p_1>p_2$.

Proof: In order to satisfy the equality ($\bigstar$), the only terms that can be negative are the ones in between squared brackets. Since $\frac{A_1}{A_2} > 1$ and $\min_i \left\{\frac{1}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}\right\} > 1$, the only way one of these terms is negative is if:

$$\min_i \left\{ \left(\frac{p_2}{p_1}\right)^{\frac{1-(1-\alpha)\lambda_{\min\{i,N-(i-1)\}_i}\gamma}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}} \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}} \right\} < 1 \tag{A.12}$$

But, since $\min_i \left\{ \left(\frac{p_2}{p_1}\right)^{\frac{1-(1-\alpha)\lambda_{\min\{i,N-(i-1)\}_i}\gamma}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}} \left(\frac{A_1}{A_2}\right)^{\frac{1}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}} \right\} > 1$, the only way that the above inequality is satisfied is if $\frac{p_2}{p_1} < 1 \Rightarrow p_2 < p_1$. $\blacksquare$

Theorem A.1: City Size and TFP. Let $A_1 > A_2$, $\lambda_i > 1, \lambda_i\gamma < 1$, for every $i \in \left\{1, 2, ..., \frac{N}{2}\right\}$ and $\{\lambda_i\}_{i=1}^{\frac{N}{2}}$ be a decreasing sequence. Then, the more productive city is larger, $S_1 > S_2$.

Proof: Based on Lemma A.1, we know that $p_1 > p_2$. Since $\lambda_i > 1, \lambda_i\gamma < 1$, for every $i \in \left\{1, 2, ..., \frac{N}{2}\right\}$ and $\{\lambda_i\}_{i=1}^{\frac{N}{2}}$ is a decreasing sequence, we know that $\frac{1}{1-\lambda_1\gamma} > \frac{1}{1-\lambda_2\gamma} > \cdots > \frac{1}{1-\lambda_{\frac{N}{2}}\gamma} > 1$ and $\frac{1-(1-\alpha)\lambda_1\gamma}{1-\lambda_1\gamma} > \frac{1-(1-\alpha)\lambda_1\gamma}{1-\lambda_1\gamma} > \cdots > \frac{1-(1-\alpha)\lambda_{\frac{N}{2}}\gamma}{1-\lambda_{\frac{N}{2}}\gamma} > 1$. But then, in order to satisfy ($\bigstar$), we must have some positive and negative terms. The term with respect to $i = \frac{N}{2}$ is positive if:

$$\frac{p_2}{p_1} > \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-(1-\alpha)\lambda_{\frac{N}{2}}\gamma}} \tag{A.13}$$

While the term with respect to $i = 1$ is positive if:

$$\frac{p_2}{p_1} > \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-(1-\alpha)\lambda_1\gamma}} \tag{A.14}$$

But notice that $\frac{A_2}{A_1} < 1$. Then, we have that:

$$\left(\frac{A_2}{A_1}\right)^{\frac{1}{1-(1-\alpha)\lambda_{\frac{N}{2}}\gamma}} > \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-(1-\alpha)\lambda_1\gamma}} \tag{A.15}$$

Therefore, in order to satisfy ($\bigstar$), we must have that:

$$\frac{p_2}{p_1} \in \left(\left(\frac{A_2}{A_1}\right)^{\frac{1}{1-(1-\alpha)\lambda_1\gamma}}, \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-(1-\alpha)\lambda_{\frac{N}{2}}\gamma}} \right) \tag{A.16}$$

But this implies that:

$$\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2} \in \left(\left(\frac{A_2}{A_1}\right)^{\frac{\alpha}{1-(1-\alpha)\lambda_1\gamma}-1}, \left(\frac{A_2}{A_1}\right)^{\frac{\alpha}{1-(1-\alpha)\lambda_{\frac{N}{2}}\gamma}-1} \right) \tag{A.17}$$

Rearranging it, we have:

$$\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2} \in \left(\left(\frac{A_1}{A_2}\right)^{\frac{(1-\alpha)(1-\lambda_1\gamma)}{1-(1-\alpha)\lambda_1\gamma}}, \left(\frac{A_1}{A_2}\right)^{\frac{(1-\alpha)\left(1-\lambda_{\frac{N}{2}}\gamma\right)}{1-(1-\alpha)\lambda_{\frac{N}{2}}\gamma}} \right) \tag{A.18}$$

Since $A_1 > A_2$, we have that:

$$\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2} > 1 \tag{A.19}$$

From the expressions for m_{ij} :

$$m_{i,1} = \frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\lambda_{\min\{i,N-(i-1)\}_i}\gamma-1}} M_i}{C_2 + \left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\lambda_{\min\{i,N-(i-1)\}_i}\gamma-1}} C_1} \tag{A.20}$$

$$m_{i2} = \frac{M_i}{C_2 + \left[\left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1}\right]^{\frac{1}{\lambda_{\min\{i,N-(i-1)\}_i}\gamma-1}} C_1} \tag{A.21}$$

and $\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2} > 1$, we have that $m_{i1} > m_{i2}$, for every $i \in \{1, 2, ..., N\}$. Finally, since:

$$S_j = \sum_{i=1}^{N} m_{ij} \tag{A.22}$$

it immediately follows that $S_1 > S_2$. ■

Theorem A.2: thick tails. Given that $A_1 > A_2$, $\lambda_i > 1, \lambda_i \gamma < 1$, for every $i \in \left\{1, 2, ..., \frac{N}{2}\right\}$ and $\{\lambda_i\}_{i=1}^{\frac{N}{2}}$ is a decreasing sequence , t;he skill distribution in the larger city has thicker tails.

Proof:

$$pdf_{11} = \frac{\dfrac{\left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_1\gamma}} M_1}{C_2 + \left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_1\gamma}} C_1}}{\sum_{i=1}^{N} \dfrac{\left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}} M_i}{C_2 + \left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}} C_1}} \tag{A.23}$$

$$= \frac{1}{\sum_{i=1}^{N} \frac{M_i}{M_1} \left\{ \begin{array}{c} \left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{\left(\lambda_{\min\{i,N-(i-1)\}_i} - \lambda_1\right)\gamma}{\left(1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma\right)(1-\lambda_1\gamma)}} \times \\ \times \dfrac{C_2 + \left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_1\gamma}} C_1}{C_2 + \left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}} C_1} \end{array} \right\}} \tag{A.24}$$

while:

$$pdf_{12} = \frac{\dfrac{M_1}{C_2 + \left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_1\gamma}} C_1}}{\sum_{i=1}^{N} \dfrac{M_i}{C_2 + \left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}} C_1}} \tag{A.25}$$

$$= \frac{1}{\sum_{i=1}^{N} \frac{M_i}{M_1} \dfrac{C_2 + \left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_1\gamma}} C_1}{C_2 + \left[\left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}} C_1}} \tag{A.26}$$

But then, since $\lambda_1 = \max\{\lambda_i\}_{i=1}^N$ and $\left(\frac{p_2}{p_1}\right)^\alpha \frac{A_1}{A_2} > 1$, we have that:

$$\sum_{i=1}^{N} \frac{M_i}{M_1} \left\{ \begin{array}{c} \left[\left(\frac{p_2}{p_1}\right)^\alpha \frac{A_1}{A_2}\right]^{\frac{\left(\lambda_{\min\{i,N-(i-1)\}_i} - \lambda_1\right)\gamma}{\left(1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma\right)(1-\lambda_1\gamma)}} \times \\ \times \frac{C_2 + \left[\left(\frac{p_2}{p_1}\right)^\alpha \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_1\gamma}} C_1}{C_2 + \left[\left(\frac{p_2}{p_1}\right)^\alpha \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}} C_1} \end{array} \right\} \tag{A.27}$$

$$< \sum_{i=1}^{N} \frac{M_i}{M_1} \frac{C_2 + \left[\left(\frac{p_2}{p_1}\right)^\alpha \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_1\gamma}} C_1}{C_2 + \left[\left(\frac{p_2}{p_1}\right)^\alpha \frac{A_1}{A_2}\right]^{\frac{1}{1-\lambda_{\min\{i,N-(i-1)\}_i}\gamma}} C_1} \tag{A.28}$$

Therefore, $pdf_{11} > pdf_{12}$. Since the distributions are symmetric, we also have $pdf_{N1} > pdf_{N2}$.■

II. Nested CES and Free Entry of firms

We now consider a technology with gross complementarities β and 3 skill types:

$$Y = A_1 \left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1 \right]^{\lambda} \right]^{\beta}. \tag{A.29}$$

In this model we simultaneously consider the additional extension that firms are perfectly mobile. Firms can relocate instantaneously and at no cost to another city. To establish itself in a city, a firm must buy a amount k of land. Given that firms can freely enter and exit cities, we have that in equilibrium, firms must generate zero profits, i.e.:

$$A_j F(m_{1j}, m_{2j}, m_{3j}) - \sum_{i}^{3} w_{ij} m_{ij} - k p_j = 0, \forall j \in \{1, 2\} \tag{A.30}$$

We will assume that there are only two cities, 1 and 2, while city i has a measure N_i of firms, that will be pin down in equilibrium. Since $w_{i2} = \left(\frac{p_2}{p_1}\right)^{\alpha} w_{i1}$, the system then becomes:

$$\left\{ \begin{array}{lr}
A_1 \dfrac{\left[\begin{array}{c} \left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1 \right]^{\lambda} \right]^{\beta-1} \times \\ \times \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1 \right]^{\lambda-1} \end{array} \right]}{\left[\begin{array}{c} \left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1 \right]^{\lambda} \right]^{\beta-1} \times \\ \times \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1 \right]^{\lambda-1} \end{array} \right]} m_{11}^{\gamma-1} = \left(\frac{p_1}{p_2}\right)^{\alpha} A_2 m_{12}^{\gamma-1} & (1) \\
A_1 \left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1 \right]^{\lambda} \right]^{\beta-1} m_{21}^{\gamma-1} = \left(\frac{p_1}{p_2}\right)^{\alpha} A_2 \left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1 \right]^{\lambda} \right]^{\beta-1} m_{22}^{\gamma-1} & (2) \\
A_1 \dfrac{\left[\begin{array}{c} \left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1 \right]^{\lambda} \right]^{\beta-1} \times \\ \times \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1 \right]^{\lambda-1} \end{array} \right]}{\left[\begin{array}{c} {}_2 \left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1 \right]^{\lambda} \right]^{\beta-1} \times \\ \times \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1 \right]^{\lambda-1} \end{array} \right]} m_{31}^{\gamma-1} = \left(\frac{p_1}{p_2}\right)^{\alpha} A_2 m_{32}^{\gamma-1} & (3) \\
N_1 m_{i1} + N_2 m_{i2} = M_i, \ \forall i \in \{1, 2, 3\} & (4, 5, 6) \\
\left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1 \right]^{\lambda} \right]^{\beta-1} \left\{ \lambda \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1 \right]^{\lambda} + m_{21}^{\gamma} y_2 \right\} = \left[\frac{H}{N_1} - k \right] \frac{p_1}{\alpha\gamma\beta A_1} & (7) \\
\left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1 \right]^{\lambda} \right]^{\beta-1} \left\{ \lambda \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1 \right]^{\lambda} + m_{22}^{\gamma} y_2 \right\} = \left[\frac{H}{N_2} - k \right] \frac{p_2}{\alpha\gamma\beta A_2} & (8) \\
\left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1 \right]^{\lambda} \right]^{\beta-1} \left\{ (1 - \lambda\gamma\beta) \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1 \right]^{\lambda} + (1 - \gamma\beta) m_{21}^{\gamma} y_2 \right\} = \frac{k}{A_1} p_1 & (9) \\
\left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1 \right]^{\lambda} \right]^{\beta-1} \left\{ (1 - \lambda\gamma\beta) \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1 \right]^{\lambda} + (1 - \gamma\beta) m_{22}^{\gamma} y_2 \right\} = \frac{k}{A_2} p_2 & (10)
\end{array} \right. \tag{A.31}$$

From eq. (1) and (3), we have:

$$\begin{aligned} \left[\frac{m_{11}}{m_{31}} \right]^{\gamma-1} &= \left[\frac{m_{12}}{m_{32}} \right]^{\gamma-1} \\ \frac{m_{11}}{m_{31}} &= \frac{m_{12}}{m_{32}} \end{aligned}$$

Since:

$$m_{12} = \frac{M_1}{N_2} - \frac{N_1}{N_2} m_{11} \tag{A.32}$$

$$m_{32} = \frac{M_3}{N_2} - \frac{N_1}{N_2} m_{31} \tag{A.33}$$

Substituting it, we have:

$$\frac{m_{11}}{m_{31}} = \frac{M_1 - N_1 m_{11}}{M_3 - N_1 m_{31}} \tag{A.34}$$

Rearranging:

$$m_{11} = \frac{M_1}{M_3} m_{31} \tag{A.35}$$

Considering a symmetric distribution (so $M_1 = M_3$), we have that $m_{11} = m_{31}$. Similarly $m_{12} = m_{32} = \frac{M_1}{N_2} - \frac{N_1}{N_2} m_{11}$.

From (1) and (2), we have:

$$\frac{\left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda-1} m_{11}^{\gamma-1}}{m_{21}^{\gamma-1}} = \frac{\left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda-1} m_{12}^{\gamma-1}}{m_{22}^{\gamma-1}} \tag{A.36}$$

Using the symmetry of the distribution and consequentially that $m_{11} = m_{31}$ and $m_{12} = m_{32}$, we have:

$$m_{11}^{\gamma(\lambda-1)} \left(\frac{m_{11}}{m_{21}}\right)^{\gamma-1} = m_{12}^{\gamma(\lambda-1)} \left(\frac{m_{12}}{m_{22}}\right)^{\gamma-1} \tag{A.37}$$

Then:

$$m_{21} = \left(\frac{m_{11}}{m_{12}}\right)^{\frac{\lambda\gamma-1}{\gamma-1}} m_{22} \tag{A.38}$$

Then, from (7) and (9), we have:

$$\frac{\left\{\lambda \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda} + m_{21}^{\gamma} y_2\right\}}{\left\{(1-\lambda\gamma\beta) \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda} + (1-\gamma\beta)\, m_{21}^{\gamma} y_2\right\}} = \frac{\left[\frac{H}{N_1} - k\right]}{k\alpha\gamma\beta} \tag{A.39}$$

Using symmetry and again that $m_{11} = m_{31}$, we have:

$$m_{21}^{\gamma} y_2 = \frac{\left\{(1-\lambda\gamma\beta)\left[\frac{H}{N_1 k\alpha\gamma\beta} - \frac{k}{k\alpha\gamma\beta}\right] - \lambda\right\}}{\left\{1 - (1-\gamma\beta)\left[\frac{H}{N_1 k\alpha\gamma\beta} - \frac{k}{k\alpha\gamma\beta}\right]\right\}} m_{11}^{\lambda\gamma} \left[y_3 + y_1\right]^{\lambda} \tag{A.40}$$

Similarly, from (8) and (10), we have:

$$\frac{\left\{\lambda \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda} + m_{22}^{\gamma} y_2\right\}}{\left\{(1-\lambda\gamma\beta) \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda} + (1-\gamma\beta)\, m_{22}^{\gamma} y_2\right\}} = \frac{\left[\frac{H}{N_2} - k\right]}{k\alpha\gamma\beta} \tag{A.41}$$

Using symmetry and again that $m_{12} = m_{32}$, we have:

$$m_{22}^{\gamma} y_2 = \frac{\left\{ (1-\lambda\gamma\beta)\left[\frac{H}{N_2 k\alpha\gamma\beta} - \frac{k}{k\alpha\gamma\beta} \right] - \lambda \right\}}{\left\{ 1 - (1-\gamma\beta)\left[\frac{H}{N_2 k\alpha\gamma\beta} - \frac{k}{k\alpha\gamma\beta} \right] \right\}} m_{12}^{\lambda\gamma} \left[y_3 + y_1 \right]^{\lambda} \tag{A.42}$$

Then, from equaiton (1), we have - again using symmetry:

$$\left[m_{21}^{\gamma} y_2 + m_{11}^{\lambda\gamma} \left[y_3 + y_1 \right]^{\lambda} \right]^{\beta-1} m_{11}^{\lambda\gamma-1} = \left(\frac{p_1}{p_2} \right)^{\alpha} \frac{A_2}{A_1} \left[m_{22}^{\gamma} y_2 + m_{12}^{\lambda\gamma} \left[y_3 + y_1 \right]^{\lambda} \right]^{\beta-1} m_{12}^{\lambda\gamma-1} \tag{A.43}$$

Substituting $m_{21}^{\gamma} y_2$ and $m_{22}^{\gamma} y_2$, we have:

$$\left[\frac{\beta\gamma(\lambda-1)\left[(1-\alpha)kN_1 - H\right]}{\left[1-(1-\alpha)\beta\gamma\right]kN_1 - (1-\beta\gamma)H} \right]^{\beta-1} m_{11}^{\lambda\gamma\beta-1} =$$
$$= \left(\frac{p_1}{p_2} \right)^{\alpha} \frac{A_2}{A_1} \left[\frac{\beta\gamma(\lambda-1)\left[(1-\alpha)kN_2 - H\right]}{\left[1-(1-\alpha)\beta\gamma\right]kN_2 - (1-\beta\gamma)H} \right]^{\beta-1} m_{12}^{\lambda\gamma\beta-1}$$

Assuming $\lambda \neq 1$, we have:

$$\left(\frac{m_{11}}{m_{12}} \right) = \left\{ \left(\frac{p_1}{p_2} \right)^{\alpha} \frac{A_2}{A_1} \left[\begin{array}{c} \frac{[1-(1-\alpha)\beta\gamma]kN_1 - (1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2 - (1-\beta\gamma)H} \times \\ \times \frac{(1-\alpha)kN_2 - H}{(1-\alpha)kN_1 - H} \end{array} \right]^{\beta-1} \right\}^{\frac{1}{\lambda\gamma\beta-1}} \tag{A.44}$$

Substituting $m_{12} = \frac{M_1}{N_2} - \frac{N_1}{N_2} m_{11}$, we have:

$$m_{11} = \frac{\left\{ \left(\frac{p_1}{p_2} \right)^{\alpha} \frac{A_2}{A_1} \left[\begin{array}{c} \frac{[1-(1-\alpha)\beta\gamma]kN_1 - (1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2 - (1-\beta\gamma)H} \\ \times \frac{(1-\alpha)kN_2 - H}{(1-\alpha)kN_1 - H} \end{array} \right]^{\beta-1} \right\}^{\frac{1}{\lambda\gamma\beta-1}}}{\left[N_2 + N_1 \left\{ \left(\frac{p_1}{p_2} \right)^{\alpha} \frac{A_2}{A_1} \left[\begin{array}{c} \frac{[1-(1-\alpha)\beta\gamma]kN_1 - (1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2 - (1-\beta\gamma)H} \\ \times \frac{(1-\alpha)kN_2 - H}{(1-\alpha)kN_1 - H} \end{array} \right]^{\beta-1} \right\}^{\frac{1}{\lambda\gamma\beta-1}} \right]} M_1 \tag{A.45}$$

Since the distribution is symmetric, we have:

$$m_{31} = \frac{\left\{ \left(\frac{p_1}{p_2} \right)^{\alpha} \frac{A_2}{A_1} \left[\begin{array}{c} \frac{[1-(1-\alpha)\beta\gamma]kN_1 - (1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2 - (1-\beta\gamma)H} \\ \times \frac{(1-\alpha)kN_2 - H}{(1-\alpha)kN_1 - H} \end{array} \right]^{\beta-1} \right\}^{\frac{1}{\lambda\gamma\beta-1}}}{\left[N_2 + N_1 \left\{ \left(\frac{p_1}{p_2} \right)^{\alpha} \frac{A_2}{A_1} \left[\begin{array}{c} \frac{[1-(1-\alpha)\beta\gamma]kN_1 - (1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2 - (1-\beta\gamma)H} \\ \times \frac{(1-\alpha)kN_2 - H}{(1-\alpha)kN_1 - H} \end{array} \right]^{\beta-1} \right\}^{\frac{1}{\lambda\gamma\beta-1}} \right]} M_3 \tag{A.46}$$

Finally, from our expression for $m_{21}^{\gamma} y_2$, we have:

$$m_{21}^{\gamma} y_2 = \frac{\left\{ \frac{(1-\lambda\gamma\beta)[H-N_1 k]-\lambda N_1 k\alpha\gamma\beta}{N_1 k\alpha\gamma\beta} \right\}}{\left\{ \frac{N_1 k\alpha\gamma\beta-(1-\gamma\beta)[H-N_1 k]}{N_1 k\alpha\gamma\beta} \right\}} m_{11}^{\lambda\gamma} \left[y_3 + y_1\right]^{\lambda} \tag{A.47}$$

Rearranging it, we have:

$$m_{21} = \left[\frac{H\left(1-\beta\lambda\gamma\right) - \left[1-\left(1-\alpha\right)\beta\lambda\gamma\right] kN_1}{\left[1-\left(1-\alpha\right)\beta\gamma\right] kN_1 - \left(1-\beta\gamma\right) H} \right]^{\frac{1}{\gamma}} m_{11}^{\lambda} \left(\frac{\left[y_3+y_1\right]^{\lambda}}{y_2} \right)^{\frac{1}{\gamma}} \tag{A.48}$$

Substituting m_{11}, we have:

$$m_{21} = \left\{ \begin{array}{c} \left[\frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_1}{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H} \right]^{\frac{1}{\gamma}} \left(\frac{[y_3+y_1]^{\lambda}}{y_2} \right)^{\frac{1}{\gamma}} \times \\ \times \frac{\left\{ \left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1} \left[\begin{array}{c} \frac{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H} \\ \times \frac{(1-\alpha)kN_2-H}{(1-\alpha)kN_1-H} \end{array} \right]^{\beta-1} \right\}^{\frac{\lambda}{\lambda\gamma\beta-1}}}{\left[N_2+N_1 \left\{ \left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1} \left[\begin{array}{c} \frac{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H} \\ \times \frac{(1-\alpha)kN_2-H}{(1-\alpha)kN_1-H} \end{array} \right]^{\beta-1} \right\}^{\frac{1}{\lambda\gamma\beta-1}} \right]^{\lambda}} \left(M_1\right)^{\lambda} \end{array} \right\} \tag{A.49}$$

Then, also notice that:

$$m_{12} = \frac{M_1}{N_2 + N_1 \left\{ \left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1} \left[\begin{array}{c} \frac{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H} \\ \times \frac{(1-\alpha)kN_2-H}{(1-\alpha)kN_1-H} \end{array} \right]^{\beta-1} \right\}^{\frac{1}{\lambda\gamma\beta-1}}} \tag{A.50}$$

and:

$$m_{32} = \frac{M_3}{\left[N_2 + N_1 \left\{ \left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1} \left[\begin{array}{c} \frac{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H} \\ \times \frac{(1-\alpha)kN_2-H}{(1-\alpha)kN_1-H} \end{array} \right]^{\beta-1} \right\}^{\frac{1}{\lambda\gamma\beta-1}} \right]} \tag{A.51}$$

and

$$m_{22} = \left\{ \begin{array}{c} \left[\frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H} \right]^{\frac{1}{\gamma}} \left(\frac{[y_3+y_1]^{\lambda}}{y_2} \right)^{\frac{1}{\gamma}} \times \\ \times \left(\frac{M_1}{N_2+N_1 \left\{ \left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1} \left[\frac{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H} \times \frac{(1-\alpha)kN_2-H}{(1-\alpha)kN_1-H} \right]^{\beta-1} \right\}^{\frac{1}{\lambda\gamma\beta-1}}} \right)^{\lambda} \end{array} \right\} \tag{A.52}$$

Proposition A. 1 *If $\lambda > 1, \lambda\gamma < 1$, and $\lambda\gamma\beta < 1$, there is no equilibrium in which $A_2 > A_1$, and $m_{i1} \geq m_{i2}$.*

Assume $A_2 > A_1$. Before we continue, we prove the following Lemma:

Lemma A. 1 *If $m_{11} > m_{12}$, then $p_1 > p_2$, $N_2 > N_1$, and $m_{21} > m_{22}$*

Proof. Going back to the system, we have:

$$\begin{cases} A_1 \dfrac{\left[\begin{array}{c} \left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \times \\ \times \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda-1} \end{array}\right]}{\left[\begin{array}{c} \left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \times \\ \times \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda-1} \end{array}\right]} m_{11}^{\gamma-1} = \left(\frac{p_1}{p_2}\right)^{\alpha} A_2 m_{12}^{\gamma-1} & (1') \\ A_1 \left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} m_{21}^{\gamma-1} = \left(\frac{p_1}{p_2}\right)^{\alpha} A_2 \left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} m_{22}^{\gamma-1} & (2') \\ A_1 \dfrac{\left[\begin{array}{c} \left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \times \\ \times \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda-1} \end{array}\right]}{\left[\begin{array}{c} \left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \times \\ \times \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda-1} \end{array}\right]} m_{31}^{\gamma-1} = \left(\frac{p_1}{p_2}\right)^{\alpha} A_2 m_{32}^{\gamma-1} & (3') \\ N_1 m_{i1} + N_2 m_{i2} = M_i, \forall i \in \{1,2,3\} & (4',5',6') \\ \left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \left\{\lambda \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda} + m_{21}^{\gamma} y_2\right\} = \left[\frac{H}{N_1} - k\right] \frac{p_1}{\alpha\gamma\beta A_1} & (7') \\ \left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \left\{\lambda \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda} + m_{22}^{\gamma} y_2\right\} = \left[\frac{H}{N_2} - k\right] \frac{p_2}{\alpha\gamma\beta A_2} & (8') \\ \left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \left\{(1-\lambda\gamma\beta) \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda} + (1-\gamma\beta) m_{21}^{\gamma} y_2\right\} = \frac{k}{A_1} p_1 & (9') \\ \left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \left\{(1-\lambda\gamma\beta) \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda} + (1-\gamma\beta) m_{22}^{\gamma} y_2\right\} = \frac{k}{A_2} p_2 & (10') \end{cases} \tag{A.53}$$

From the last two equations, we have:

$$\begin{array}{l} \left[\begin{array}{c} \left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \times \\ \times \left\{(1-\lambda\gamma\beta) \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda} + (1-\gamma\beta) m_{21}^{\gamma} y_2\right\} \end{array}\right] \frac{p_2}{p_1} \frac{A_1}{A_2} = \frac{k}{A_2} p_2 \\ \left[\begin{array}{c} \left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \times \\ \times \left\{(1-\lambda\gamma\beta) \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda} + (1-\gamma\beta) m_{22}^{\gamma} y_2\right\} \end{array}\right] = \frac{k}{A_2} p_2 \end{array} \tag{A.54}$$

Equating this two expressions, we have:

$$\left[\begin{array}{c} \left[\begin{array}{c} \left[m_{21}^{\gamma} y_2 + \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \times \\ \times \left\{(1-\lambda\gamma\beta) \left[m_{31}^{\gamma} y_3 + m_{11}^{\gamma} y_1\right]^{\lambda} + (1-\gamma\beta) m_{21}^{\gamma} y_2\right\} \end{array}\right] \frac{p_2}{p_1} \frac{A_1}{A_2} \\ - \left[\begin{array}{c} \left[m_{22}^{\gamma} y_2 + \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda}\right]^{\beta-1} \times \\ \times \left\{(1-\lambda\gamma\beta) \left[m_{32}^{\gamma} y_3 + m_{12}^{\gamma} y_1\right]^{\lambda} + (1-\gamma\beta) m_{22}^{\gamma} y_2\right\} \end{array}\right] \end{array}\right] = 0 \tag{A.55}$$

since $M_1 = M_3$, we have that $m_{31} = m_{11}$ and $m_{32} = m_{12}$. Based on these results, we have:

$$\left[\begin{array}{c} \left[m_{21}^{\gamma} y_2 + m_{11}^{\lambda\gamma} \left[y_3 + y_1\right]^{\lambda}\right]^{\beta-1} \left\{(1-\lambda\gamma\beta) m_{11}^{\lambda\gamma} \left[y_3 + y_1\right]^{\lambda} + (1-\gamma\beta) m_{21}^{\gamma} y_2\right\} \frac{p_2}{p_1} \frac{A_1}{A_2} \\ - \left[m_{22}^{\gamma} y_2 + m_{12}^{\lambda\gamma} \left[y_3 + y_1\right]^{\lambda}\right]^{\beta-1} \left\{(1-\lambda\gamma\beta) m_{12}^{\lambda\gamma} \left[y_3 + y_1\right]^{\lambda} + (1-\gamma\beta) m_{22}^{\gamma} y_2\right\} \end{array}\right] = 0 \tag{A.56}$$

Then, from equation (1), we have - again using symmetry:

$$\left[m_{21}^{\gamma}y_2+m_{11}^{\lambda\gamma}\left[y_3+y_1\right]^{\lambda}\right]^{\beta-1}=\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}\left[m_{22}^{\gamma}y_2+m_{12}^{\lambda\gamma}\left[y_3+y_1\right]^{\lambda}\right]^{\beta-1}\frac{m_{12}^{\lambda\gamma-1}}{m_{11}^{\lambda\gamma-1}} \tag{A.57}$$

Substituting it back, we have:

$$\left[m_{22}^{\gamma}y_2+m_{12}^{\lambda\gamma}\left[y_3+y_1\right]^{\lambda}\right]^{\beta-1}\left[\begin{array}{c}(1-\lambda\gamma\beta)\left[y_3+y_1\right]^{\lambda}\left[\left(\frac{p_2}{p_1}\right)^{1-\alpha}m_{11}-m_{12}\right]m_{12}^{\lambda\gamma-1}\\ +(1-\gamma\beta)\,y_2\left[\left(\frac{p_2}{p_1}\right)^{1-\alpha}\left(\frac{m_{12}}{m_{11}}\right)^{\lambda\gamma-1}m_{21}^{\gamma}-m_{22}^{\gamma}\right]\end{array}\right]=0 \qquad (\bigstar) \tag{A.58}$$

Since:

$$\begin{aligned}\left(\frac{m_{11}}{m_{12}}\right)^{\lambda\gamma-1} &= \left(\frac{m_{21}}{m_{22}}\right)^{\gamma-1}\\ &\Downarrow\\ \left(\frac{m_{12}}{m_{11}}\right)^{\lambda\gamma-1} &= \left(\frac{m_{22}}{m_{21}}\right)^{\gamma-1}\end{aligned}$$

we have:

$$\left[m_{22}^{\gamma}y_2+m_{12}^{\lambda\gamma}\left[y_3+y_1\right]^{\lambda}\right]^{\beta-1}\left\{\begin{array}{c}(1-\lambda\gamma\beta)\left[y_3+y_1\right]^{\lambda}\left[\left(\frac{p_2}{p_1}\right)^{1-\alpha}\frac{m_{11}}{m_{12}}-1\right]m_{12}^{\lambda\gamma}\\ +(1-\gamma\beta)\,y_2\left[\left(\frac{p_2}{p_1}\right)^{1-\alpha}\frac{m_{21}}{m_{22}}-1\right]m_{22}^{\gamma}\end{array}\right\}=0 \tag{A.59}$$

Since:

$$\frac{m_{21}}{m_{22}}=\left(\frac{m_{11}}{m_{12}}\right)^{\frac{\lambda\gamma-1}{\gamma-1}} \tag{A.60}$$

we have:

$$\left[m_{22}^{\gamma}y_2+m_{12}^{\lambda\gamma}\left[y_3+y_1\right]^{\lambda}\right]^{\beta-1}\left\{\begin{array}{c}(1-\lambda\gamma\beta)\left[y_3+y_1\right]^{\lambda}\left[\left(\frac{p_2}{p_1}\right)^{1-\alpha}\frac{m_{11}}{m_{12}}-1\right]m_{12}^{\lambda\gamma}\\ (1-\gamma\beta)\,y_2\left[\left(\frac{p_2}{p_1}\right)^{1-\alpha}\left(\frac{m_{11}}{m_{12}}\right)^{\frac{\lambda\gamma-1}{\gamma-1}}-1\right]m_{22}^{\gamma}\end{array}\right\}=0 \tag{A.61}$$

Assuming that $\lambda\gamma\beta<1$, we have that the only terms that can be negative are the ones inside the squared brackets inside the curly brackets.

Since $\frac{\lambda\gamma-1}{\gamma-1}\in(0,1)$:

$$\begin{aligned}&\left[\left(\frac{p_2}{p_1}\right)^{1-\alpha}\frac{m_{11}}{m_{12}}-1\right]-\left[\left(\frac{p_2}{p_1}\right)^{1-\alpha}\left(\frac{m_{11}}{m_{12}}\right)^{\frac{\lambda\gamma-1}{\gamma-1}}-1\right]=\\ &=\left(\frac{p_2}{p_1}\right)^{1-\alpha}\frac{m_{11}}{m_{12}}\left[1-\left(\frac{m_{12}}{m_{11}}\right)^{\frac{\gamma(\lambda-1)}{1-\gamma}}\right]\end{aligned}$$

Since $\frac{\gamma(\lambda-1)}{1-\gamma} \in (0,1)$, the sign will depend on $\frac{m_{12}}{m_{11}}$. Since

$$\frac{m_{12}}{m_{11}} < 1 \Rightarrow \left[\left(\frac{p_2}{p_1}\right)^{1-\alpha} \frac{m_{11}}{m_{12}} - 1\right] > \left[\left(\frac{p_2}{p_1}\right)^{1-\alpha} \left(\frac{m_{11}}{m_{12}}\right)^{\frac{\lambda\gamma-1}{\gamma-1}} - 1\right].$$

In order to keep the equality, we must have:

$$\left(\frac{p_2}{p_1}\right)^{1-\alpha} \frac{m_{11}}{m_{12}} - 1 > 0$$
$$\frac{m_{11}}{m_{12}} > \left(\frac{p_1}{p_2}\right)^{1-\alpha}$$

and

$$\left(\frac{p_2}{p_1}\right)^{1-\alpha} \left(\frac{m_{11}}{m_{12}}\right)^{\frac{\lambda\gamma-1}{\gamma-1}} - 1 < 0$$
$$\left(\frac{m_{11}}{m_{12}}\right)^{\frac{1-\lambda\gamma}{1-\gamma}} < \left(\frac{p_1}{p_2}\right)^{1-\alpha}$$

since $\frac{1-\lambda\gamma}{1-\gamma} \in (0,1)$ and $\alpha \in (0,1)$, we have that:

$$\left(\frac{p_1}{p_2}\right)^{1-\alpha} \in \left(\left(\frac{m_{11}}{m_{12}}\right)^{\frac{1-\lambda\gamma}{1-\gamma}}, \frac{m_{11}}{m_{12}}\right) \tag{A.62}$$

since $\frac{m_{11}}{m_{12}} > 1$, we have that $p_1 > p_2$.

We also showed earlier that:

$$\frac{m_{21}}{m_{22}} = \left(\frac{m_{11}}{m_{12}}\right)^{\frac{1-\lambda\gamma}{1-\gamma}} \tag{A.63}$$

since $\frac{m_{11}}{m_{12}} > 1$, we have that $m_{21} > m_{22}$.

Finally, from equations (7′) and (8′), we have:

$$\begin{aligned} N_1 &= \frac{H}{\left\{\frac{\alpha\gamma\beta A_1}{p_1}\left[m_{21}^{\gamma}y_2+\left[m_{31}^{\gamma}y_3+m_{11}^{\gamma}y_1\right]^{\lambda}\right]^{\beta-1}\left\{\lambda\left[m_{31}^{\gamma}y_3+m_{11}^{\gamma}y_1\right]^{\lambda}+m_{21}^{\gamma}y_2\right\}+k\right\}} \\ N_2 &= \frac{H}{\left\{\frac{\alpha\gamma\beta A_2}{p_2}\left[m_{22}^{\gamma}y_2+\left[m_{32}^{\gamma}y_3+m_{12}^{\gamma}y_1\right]^{\lambda}\right]^{\beta-1}\left\{\lambda\left[m_{32}^{\gamma}y_3+m_{12}^{\gamma}y_1\right]^{\lambda}+m_{22}^{\gamma}y_2\right\}+k\right\}} \end{aligned} \tag{A.64}$$

Since, from (9′) and (10′):

$$\begin{aligned} \frac{A_1}{p_1} &= \frac{k}{\left[\begin{array}{c}\left[m_{21}^{\gamma}y_2+\left[m_{31}^{\gamma}y_3+m_{11}^{\gamma}y_1\right]^{\lambda}\right]^{\beta-1}\times \\ \times\left\{(1-\beta\gamma)\,m_{21}^{\gamma}y_2+(1-\lambda\gamma\beta)\left[m_{31}^{\gamma}y_3+m_{11}^{\gamma}y_1\right]^{\lambda}\right\}\end{array}\right]} \\ \frac{A_2}{p_2} &= \frac{k}{\left[\begin{array}{c}\left[m_{22}^{\gamma}y_2+\left[m_{32}^{\gamma}y_3+m_{12}^{\gamma}y_1\right]^{\lambda}\right]^{\beta-1}\times \\ \times\left\{(1-\beta\gamma)\,m_{22}^{\gamma}y_2+(1-\lambda\gamma\beta)\left[m_{32}^{\gamma}y_3+m_{12}^{\gamma}y_1\right]^{\lambda}\right\}\end{array}\right]} \end{aligned}$$

Substituting it back, we have:

$$N_1 = \frac{H}{k}\frac{(1-\beta\gamma)\,m_{21}^{\gamma}y_2 + (1-\lambda\gamma\beta)\,m_{11}^{\lambda\gamma}\,[y_3+y_1]^{\lambda}}{[1-(1-\alpha)\,\beta\gamma]\,m_{21}^{\gamma}y_2 + [1-(1-\alpha)\,\lambda\gamma\beta]\,m_{11}^{\lambda\gamma}\,[y_3+y_1]^{\lambda}} \tag{A.65}$$

and

$$N_2 = \frac{H}{k}\frac{(1-\beta\gamma)\,m_{22}^{\gamma}y_2 + (1-\lambda\gamma\beta)\,m_{12}^{\lambda\gamma}\,[y_3+y_1]^{\lambda}}{[1-(1-\alpha)\,\beta\gamma]\,m_{22}^{\gamma}y_2 + [1-(1-\alpha)\,\lambda\gamma\beta]\,m_{12}^{\lambda\gamma}\,[y_3+y_1]^{\lambda}} \tag{A.66}$$

Then:

$$\frac{N_1}{N_2} = 1 + \frac{\alpha\beta\gamma y_2\,[y_3+y_1]^{\lambda}\,(\lambda-1)\left[\left(\frac{m_{12}}{m_{11}}\right)^{\frac{(\lambda-1)\gamma}{1-\gamma}} - 1\right] m_{22}^{\gamma}m_{11}^{\lambda\gamma}}{\left[\begin{array}{c} [1-(1-\alpha)\,\beta\gamma]\,(1-\beta\gamma)\,(y_2)^2\,m_{22}^{\gamma}m_{21}^{\gamma} \\ +[1-(1-\alpha)\,\beta\gamma]\,(1-\lambda\gamma\beta)\,y_2\,[y_3+y_1]^{\lambda}\,m_{21}^{\gamma}m_{12}^{\lambda\gamma} \\ +[1-(1-\alpha)\,\lambda\gamma\beta]\,(1-\beta\gamma)\,[y_3+y_1]^{\lambda}\,y_2 m_{22}^{\gamma}m_{11}^{\lambda\gamma} \\ +[1-(1-\alpha)\,\lambda\gamma\beta]\,(1-\lambda\gamma\beta)\,[y_3+y_1]^{2\lambda}\,m_{11}^{\lambda\gamma}m_{12}^{\lambda\gamma} \end{array}\right]} \qquad (\bigstar\bigstar) \tag{A.67}$$

Since $\frac{m_{12}}{m_{11}} < 1$, we have that $\frac{N_1}{N_2} < 1 \Rightarrow N_1 < N_2$.

Then, back in the system, rearranging it, we have:

$$\left\{\begin{array}{ll}
\left\{\begin{array}{c} \left(\frac{M_1}{N_2+N_1\left\{\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}Z^{\beta-1}\right\}^{\frac{1}{\lambda\gamma\beta-1}}}\right)^{\lambda}\left(\frac{[y_3+y_1]^{\lambda}}{y_2}\right)^{\frac{1}{\gamma}}\times \\ \times\left[\begin{array}{c} N_1\left\{\begin{array}{c}\left[\frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_1}{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}\right]^{\frac{1}{\gamma}}\times \\ \times\left\{\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}Z^{\beta-1}\right\}^{\frac{\lambda}{\lambda\gamma\beta-1}}\end{array}\right\} \\ +N_2\left[\frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H}\right]^{\frac{1}{\gamma}} \end{array}\right] \end{array}\right\} = M_2 & (1'') \\
\left[\begin{array}{c}\left\{\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}Z^{\beta-1}\right\}^{\frac{\lambda-1}{1-\lambda\gamma\beta}}\times \\ \times Z^{\frac{\gamma-1}{\gamma}}\end{array}\right] = 1 & (2'') \\
\left[\begin{array}{c}\left[\frac{[H-(1-\alpha)kN_1]\beta\gamma(\lambda-1)}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1}\right]^{\beta-1}\left[\frac{(\lambda-1)}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1}\right]\times \\ \times\left(\frac{\left\{\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}Z^{\beta-1}\right\}^{\frac{1}{\lambda\gamma\beta-1}}}{\left[N_2+N_1\left\{\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}Z^{\beta-1}\right\}^{\frac{1}{\lambda\gamma\beta-1}}\right]}M_1\right)^{\lambda\gamma\beta}[y_3+y_1]^{\lambda\beta}\end{array}\right] = \frac{1}{N_1}\frac{p_1}{\alpha\gamma\beta A_1} & (3'') \\
\left(\begin{array}{c}\left[\frac{(\lambda-1)}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}\right]\left[\frac{[H-(1-\alpha)kN_2]\beta\gamma(\lambda-1)}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}\right]^{\beta-1}\times \\ \times\left(\frac{M_1}{N_2+N_1\left\{\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A_2}{A_1}Z^{\beta-1}\right\}^{\frac{1}{\lambda\gamma\beta-1}}}\right)^{\lambda\gamma\beta}[y_3+y_1]^{\lambda\beta}\end{array}\right) = \frac{1}{N_2}\frac{p_2}{\alpha\gamma\beta A_2} & (4'')
\end{array}\right. \tag{A.68}$$

where $Z = \frac{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H} \times \frac{(1-\alpha)kN_2-H}{(1-\alpha)kN_1-H}$

Then, from $(3'')$ and $(4'')$, we have:

$$\left\{\begin{array}{c}\left(\frac{m_{11}}{m_{12}}\right)^{\lambda\gamma\beta}\left[\frac{H-(1-\alpha)kN_1}{H-(1-\alpha)kN_2}\right]^{\beta-1} \\ \times\left[\frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1}\right]^{\beta}\end{array}\right\}\frac{N_1}{N_2} = \frac{A_2}{A_1}\frac{p_1}{p_2} \tag{A.69}$$

once

$$\frac{m_{11}}{m_{12}} = \left\{ \begin{array}{c} \left(\frac{p_2}{p_1}\right)^{\alpha} \frac{A_1}{A_2} \left[\frac{(1-\beta\gamma)H-[1-(1-\alpha)\beta\gamma]kN_2}{(1-\beta\gamma)H-[1-(1-\alpha)\beta\gamma]kN_1}\right]^{\beta-1} \times \\ \times \left[\frac{H-(1-\alpha)kN_1}{H-(1-\alpha)kN_2}\right]^{\beta-1} \end{array} \right\}^{\frac{1}{1-\lambda\gamma\beta}} \tag{A.70}$$

Substituting it back, we have:

$$\left[\begin{array}{c} \left[\begin{array}{c} \frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1} \\ \times\frac{H-(1-\alpha)kN_1}{H-(1-\alpha)kN_2} \end{array} \right]^{\frac{\beta(1-\lambda\gamma)}{1-\lambda\gamma\beta}} \\ \times\left[\frac{H-(1-\alpha)kN_2}{H-(1-\alpha)kN_1}\right] \end{array} \right] \frac{N_1}{N_2} = \left(\frac{A_2}{A_1}\right)^{\frac{1}{1-\lambda\gamma\beta}} \left(\frac{p_1}{p_2}\right)^{\frac{1-(1-\alpha)\lambda\gamma\beta}{1-\lambda\gamma\beta}} \quad (\bigstar\bigstar\bigstar) \tag{A.71}$$

Notice that:

$$\left[\begin{array}{c} \frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1}\times \\ \times\frac{H-(1-\alpha)kN_1}{H-(1-\alpha)kN_2} \end{array} \right] - 1 = \frac{Hk\alpha\left(N_1 - N_2\right)}{\left\{ \begin{array}{c} \left[H\left(1-\beta\gamma\right) - \left[1-\left(1-\alpha\right)\beta\gamma\right]kN_1\right] \times \\ \times\left[H-\left(1-\alpha\right)kN_2\right] \end{array} \right\}} \tag{A.72}$$

Since:

$$m_{21} = \left[-\frac{H\left(1-\beta\lambda\gamma\right) - \left[1-\left(1-\alpha\right)\beta\lambda\gamma\right]kN_1}{H\left(1-\beta\gamma\right) - \left[1-\left(1-\alpha\right)\beta\gamma\right]kN_1}\right]^{\frac{1}{\gamma}} m_{11}^{\lambda} \left(\frac{\left[y_3+y_1\right]^{\lambda}}{y_2}\right)^{\frac{1}{\gamma}} \tag{A.73}$$

and $m_{21} > 0$, we must have that:

$$-\frac{H\left(1-\beta\lambda\gamma\right) - \left[1-\left(1-\alpha\right)\beta\lambda\gamma\right]kN_1}{H\left(1-\beta\gamma\right) - \left[1-\left(1-\alpha\right)\beta\gamma\right]kN_1} > 0 \tag{A.74}$$

Since $H\left(1-\beta\lambda\gamma\right) - \left[1-\left(1-\alpha\right)\beta\lambda\gamma\right]kN_1$ is decreasing in λ, we must have:

$$H\left(1-\beta\gamma\right) - \left[1-\left(1-\alpha\right)\beta\gamma\right]kN_1 > 0 \tag{A.75}$$

and

$$H\left(1-\beta\lambda\gamma\right) - \left[1-\left(1-\alpha\right)\beta\lambda\gamma\right]kN_1 < 0 \tag{A.76}$$

but them, we have that:

$$\left\{ \begin{array}{c} \left[H\left(1-\beta\gamma\right) - \left[1-\left(1-\alpha\right)\beta\gamma\right]kN_1\right] \times \\ \times\left[H-\left(1-\alpha\right)kN_2\right] \end{array} \right\} > 0 \tag{A.77}$$

Since $N_1 < N_2$, this implies that:

$$\left[\begin{array}{c} \frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1} \\ \times\frac{H-(1-\alpha)kN_1}{H-(1-\alpha)kN_2} \end{array} \right] < 1 \tag{A.78}$$

Then, from (★★★), we have:

$$
\begin{aligned}
\left[\begin{array}{c}\frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1} \\ \times\frac{H-(1-\alpha)kN_1}{H-(1-\alpha)kN_2}\end{array}\right]^{\frac{\beta(1-\lambda\gamma)}{1-\lambda\gamma\beta}} &< 1 \\
\left[\frac{H-(1-\alpha)\,kN_2}{H-(1-\alpha)\,kN_1}\right] &< 1 \\
\frac{N_1}{N_2} &< 1
\end{aligned}
$$

Therefore $LHS < 1$. We also know that $\left(\frac{p_1}{p_2}\right)^{\frac{1-(1-\alpha)\lambda\gamma\beta}{1-\lambda\gamma\beta}} > 1$. Given $A_2 > A_1$, $RHS > 1$ and we have a contradiction

In order to complete our proof, assume that $\frac{m_{11}}{m_{12}} = 1 \Rightarrow m_{11} = m_{12}$. Given that $\frac{m_{21}}{m_{22}} = \left(\frac{m_{11}}{m_{12}}\right)^{\frac{1-\lambda\gamma}{1-\gamma}}$, we have that $m_{21} = m_{22}$. Then, from (★★) we have $N_1 = N_2$ and from (★) we have $p_1 = p_2$. But, combining these results and (★★★), we again have a contradiction, once $LHS = 1$ while RHS¿1 once $A_2 > A_1$. ■

Corollary A. 1 *There is no equilibrium in which $A_1 > A_2$ and $m_{i2} > m_{i1}$, $\forall i \in \{1, 2, 3\}$.*

Theorem A. 1 *City Size and TFP. Let $A_1 > A_2, \beta > 1, \lambda\gamma\beta < 1$, and $\gamma < 1$. Then the more productive city is larger, $S_1 > S_2$.*

Proof. Before we start, define:

$$
Z = \left(\frac{p_1}{p_2}\right)^{\alpha} \frac{A_2}{A_1} \left[\begin{array}{c}\frac{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H} \\ \times\frac{(1-\alpha)kN_2-H}{(1-\alpha)kN_1-H}\end{array}\right]^{\beta-1} \tag{A.79}
$$

Notice that:

$$
\frac{m_{11}}{m_{12}} = Z^{\frac{1}{\lambda\gamma\beta-1}} \tag{A.80}
$$

Since $A_1 > A_2$, from Corollary 1 we have $m_{11} > m_{12}$. From Lemma 1 and $\lambda\gamma\beta < 1$, we have that $Z < 1$.

Notice that:

$$
\begin{aligned}
S_1 &= N_1\left(2 * m_{11} + m_{21}\right) \\
S_1 &= N_2 \left(\begin{array}{c} 2*\frac{N_1}{N_2}\frac{Z^{\frac{1}{\lambda\gamma\beta-1}}}{\left[N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}\right]}M_1 \\ +\frac{N_1}{N_2}\left\{\begin{array}{c}\left[\frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_1}{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}\right]^{\frac{1}{\gamma}} \\ \times\left(\frac{[y_3+y_1]^{\lambda}}{y_2}\right)^{\frac{1}{\gamma}}\left(\frac{Z^{\frac{1}{\lambda\gamma\beta-1}}}{\left[N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}\right]}M_1\right)^{\lambda}\end{array}\right\}\end{array}\right)
\end{aligned}
$$

while:

$$
\begin{aligned}
S_2 &= N_2\left(2*m_{12}+m_{22}\right) \\
&= N_2\left(\begin{array}{c} 2*\frac{M_1}{N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}} \\ +\left\{\left[\frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H}\right]^{\frac{1}{\gamma}}\left(\frac{[y_3+y_1]^{\lambda}}{y_2}\right)^{\frac{1}{\gamma}}\left(\frac{M_1}{N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}}\right)^{\lambda}\right\}\end{array}\right)
\end{aligned}
$$

then:

$$
S_2-S_1=N_2\left\{\begin{array}{c} 2\frac{\left[1-\frac{N_1}{N_2}Z^{\frac{1}{\lambda\gamma\beta-1}}\right]M_1}{N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}}+ \\ +\left[\begin{array}{c}\left[1-\frac{N_1}{N_2}Z^{\frac{\lambda}{\lambda\gamma\beta-1}}\right]\left[\frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H}\right]^{\frac{1}{\gamma}}\times \\ \times\left(\frac{[y_3+y_1]^{\lambda}}{y_2}\right)^{\frac{1}{\gamma}}\left(\frac{M_1}{N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}}\right)^{\lambda}\end{array}\right]\end{array}\right\} \tag{A.81}
$$

Since:

$$
\left\{\begin{array}{c}\left[\frac{H-(1-\alpha)kN_1}{H-(1-\alpha)kN_2}\right]^{\beta-1} \\ \times\left[\frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1}\right]^{\beta}\end{array}\right\}Z^{\frac{\lambda\gamma\beta}{\lambda\gamma\beta-1}}=\frac{N_2}{N_1}\frac{p_1}{p_2}\frac{A_2}{A_1} \tag{A.82}
$$

Then, from $\frac{(3'')}{(4'')}$, we have:

$$
\frac{N_1}{N_2}=\frac{\frac{p_1}{p_2}\frac{A_2}{A_1}}{\left[\frac{H-(1-\alpha)kN_1}{H-(1-\alpha)kN_2}\right]^{\beta-1}\left[\frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1}\right]^{\beta}Z^{\frac{\lambda\gamma\beta}{\lambda\gamma\beta-1}}} \tag{A.83}
$$

so:

$$
\begin{aligned}
\frac{N_1}{N_2}Z^{\frac{1}{\lambda\gamma\beta-1}} &= \frac{\frac{p_1}{p_2}\frac{A_2}{A_1}}{\left[\frac{H-(1-\alpha)kN_1}{H-(1-\alpha)kN_2}\right]^{\beta-1}\left[\frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1}\right]^{\beta}}\left(\frac{1}{Z}\right) \\
&= \left(\frac{p_1}{p_2}\right)^{1-\alpha}\left[\frac{H\left(1-\beta\gamma\right)-\left[1-\left(1-\alpha\right)\beta\gamma\right]kN_1}{H\left(1-\beta\gamma\right)-\left[1-\left(1-\alpha\right)\beta\gamma\right]kN_2}\right]>1
\end{aligned}
$$

Therefore, since $\frac{N_1}{N_2}Z^{\frac{1}{\lambda\gamma\beta-1}}>1$, we have that $\left[1-\frac{N_1}{N_2}Z^{\frac{1}{\lambda\gamma\beta-1}}\right]<0$. Since $\lambda>1$ and $\frac{N_1}{N_2}Z^{\frac{\lambda}{\lambda\gamma\beta-1}}.\frac{N_1}{N_2}Z^{\frac{1}{\lambda\gamma\beta-1}}$, we also have that $\left[1-\frac{N_1}{N_2}Z^{\frac{\lambda}{\lambda\gamma\beta-1}}\right]<0$. Therefore:

$$
S_2-S_1<0 \tag{A.84}
$$

and we have that the city with the highest TFP is also the largest city. ■

Theorem A. 2 *Thick tails. Let $A_1>A_2,\beta>1,\lambda>1$, and $\lambda\gamma\beta<1$, the skill distribution in the larger city has thicker tails.*

Proof. Consider the distributions, denoted by pdf_{ij} :

$$pdf_{11} = \frac{N_1 m_{11}}{S_1} = \frac{\frac{Z^{\frac{1}{\lambda\gamma\beta-1}}}{\left[N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}\right]} M_1}{\left(\begin{array}{c} 2 * \frac{Z^{\frac{1}{\lambda\gamma\beta-1}}}{\left[N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}\right]} M_1 \\ + \left\{ \begin{array}{c} \left[\frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_1}{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}\right]^{\frac{1}{\gamma}} \\ \times \left(\frac{[y_3+y_1]^\lambda}{y_2}\right)^{\frac{1}{\gamma}} \left(\frac{Z^{\frac{1}{\lambda\gamma\beta-1}}}{\left[N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}\right]} M_1\right)^\lambda \end{array} \right\} \end{array} \right)} \tag{A.85}$$

$$pdf_{12} = \frac{\frac{M_1}{N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}}}{\left(\begin{array}{c} 2 * \frac{M_1}{N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}} \\ + \left\{ \begin{array}{c} \left[\frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H}\right]^{\frac{1}{\gamma}} \\ \times \left(\frac{[y_3+y_1]^\lambda}{y_2}\right)^{\frac{1}{\gamma}} \left(\frac{M_1}{N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}}\right)^\lambda \end{array} \right\} \end{array} \right)} \tag{A.86}$$

Since:

$$\begin{aligned} & \left(\frac{M_1}{N_2 + N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}}\right)^\lambda \left(\frac{[y_3+y_1]^\lambda}{y_2}\right)^{\frac{1}{\gamma}} \\ = \quad & \frac{M_2}{\left[\begin{array}{c} N_1 \left[\frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_1}{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}\right]^{\frac{1}{\gamma}} Z^{\frac{\lambda}{\lambda\gamma\beta-1}} \\ +N_2 \left[\frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H}\right]^{\frac{1}{\gamma}} \end{array} \right]} \end{aligned}$$

we have:

$$pdf_{11} = \frac{Z^{\frac{1}{\lambda\gamma\beta-1}} M_1}{\left(\begin{array}{c} 2Z^{\frac{1}{\lambda\gamma\beta-1}} M_1 \\ + \frac{M_2\left[N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}\right]}{\left[N_1 Z^{\frac{\lambda}{\lambda\gamma\beta-1}} + N_2 \left[\begin{array}{c} \frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2}{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H} \\ \times \frac{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H}{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_1} \end{array} \right]^{\frac{1}{\gamma}} \right]} \end{array} \right)} \tag{A.87}$$

and

$$pdf_{12} = \frac{M_1}{\left(\begin{array}{c} 2M_1 \\ + \frac{M_2\left[N_2+N_1 Z^{\frac{1}{\lambda\gamma\beta-1}}\right]}{\left[N_1 Z^{\frac{\lambda}{\lambda\gamma\beta-1}} \left[\begin{array}{c} \frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_1}{[1-(1-\alpha)\beta\gamma]kN_1-(1-\beta\gamma)H} \\ \times \frac{[1-(1-\alpha)\beta\gamma]kN_2-(1-\beta\gamma)H}{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2} \end{array} \right]^{\frac{1}{\gamma}} + N_2 \right]} \end{array} \right)} \tag{A.88}$$

$$\frac{pdf_{11}}{pdf_{12}} = \left(\frac{\left\{ \begin{array}{c} \left[\begin{array}{c} 2Z^{\frac{1}{\lambda\gamma\beta-1}} M_1 \times \left[\begin{array}{c} \frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2}{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_1} \\ \times\frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2} \end{array} \right]^{\frac{1}{\gamma}} \times \\ \times \left[N_1 Z^{\frac{\lambda}{\lambda\gamma\beta-1}} \left[\begin{array}{c} \frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_1}{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2} \\ \times\frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1} \end{array} \right]^{\frac{1}{\gamma}} + N_2 \right] \end{array} \right] \\ +M_2 \left[N_2 + N_1 Z^{\frac{1}{\lambda\gamma\beta-1}} \right] Z^{\frac{1}{\lambda\gamma\beta-1}} \end{array} \right\}}{\left\{ \begin{array}{c} \left[\begin{array}{c} 2Z^{\frac{1}{\lambda\gamma\beta-1}} M_1 \times \left[\begin{array}{c} \frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2}{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_1} \\ \times\frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2} \end{array} \right]^{\frac{1}{\gamma}} \times \\ \times \left[N_1 Z^{\frac{\lambda}{\lambda\gamma\beta-1}} \left[\begin{array}{c} \frac{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_1}{H(1-\beta\lambda\gamma)-[1-(1-\alpha)\beta\lambda\gamma]kN_2} \\ \times\frac{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_2}{H(1-\beta\gamma)-[1-(1-\alpha)\beta\gamma]kN_1} \end{array} \right]^{\frac{1}{\gamma}} + N_2 \right] \end{array} \right] \\ +M_2 \left[N_2 + N_1 Z^{\frac{1}{\lambda\gamma\beta-1}} \right] \end{array} \right\}} \right) \tag{A.89}$$

Since $Z < 1 \Rightarrow Z^{\frac{1}{\lambda\gamma\beta-1}} > 1$, we have that $\frac{pdf_{11}}{pdf_{12}} > 1$. Since the distribution is symmetric, we also have that $\frac{pdf_{31}}{pdf_{32}} > 1$ ■

III. Agglomeration Externalities

Since Marshall (1890), there is a broad consensus in the economics literature that the principal explanation for the existence of cities is the presence of agglomeration externalities (see Duranton and Puga, 2004 for a theoretical survey, and Rosenthal and Strange, 2004, and Combes, Duranton, Gobillon, Puga, and Roux, 2009 for the empirical evidence). Due to economies of scale and network effects, firms that cluster together may see a decline in the costs of production, due to the presence of competing multiple suppliers, greater specialization and the division of labor. Even direct competitors in the same sector may benefit because the cluster of firms in the same city attracts more suppliers and customers than a single firm in isolation. Alternatively, there may be demand-side agglomeration externalities due to the variety of goods and services provided. Of course, the size of the city is limited by diseconomies of agglomeration, for example congestion, and the limited availability of land which drives up the price of housing and office space. The latter is captured by the fixed amount of land in our baseline model above.

Because the key feature of agglomeration externalities is city size, we will assume that TFP is determined *endogenously* and increasing in city size $S_j : A_j = A(S_j)$, where $A'(S) > 0$. Cities are ex ante identical and there are no initial differences in TFP across different cities.

We analyze the case of Extreme-Skill Complementarities with free mobility of firms in a two-city model. The next Theorem characterizes the equilibrium allocation of skilled workers and of firms across cities. Firms in larger cities are more productive due to higher agglomeration externalities. Also real estate prices are higher. This may seem obvious, but as in the results on city size with exogenous TFP, this hinges on the fact that land supply is equal (or at least not too different). Once endogenous TFP due to agglomeration is higher, labor productivity is higher and therefore labor demand. As a result, the representative firm is larger and more firms enter. Finally, the same logic that explains the emergence of thick tails applies exactly as it does for the case of exogenous TFP.

Theorem A. 3 *Given endogenous agglomeration externalities and given $\lambda > 1$ and $\lambda\gamma < 1$, and provided cities of different size exist, the larger city has:*

- *higher TFP;*
- *higher real estate prices;*
- *more and larger firms;*
- *thicker tails in the skill distribution.*

Proof. See below. ■

Observe that with endogenous agglomeration externalities we can readily extend the proofs to the case of the technology with Top-Skill Complementarity.

An open question remains whether there actually exist equilibria with endogenous agglomeration externalities where cities are ex post heterogeneous, despite being ex ante identical. Of course, if there is heterogeneity in equilibrium city size, we expect there to be multiple equilibria since there is no ex ante advantage to any city ex ante: one equilibrium where city 1 is large and city 2 is small, another equilibrium where city 2 is large and city 1 is small, and finally an intermediate equilibrium where cities are identical.

For the case of a CES production technology, we show conditions under which cities are different in size, despite the ex ante identical technologies and agglomeration externalities. Mobility of workers and free entry of firms induces wages and housing prices to adjust such that workers are indifferent between locating in either city. When there are sufficiently large economies of scale of agglomeration, i.e. when the function A is sufficiently convex, we obtain that cities differ in equilibrium. We establish this result for the exponential function in conjunction with the CES technology.

Theorem A. 4 *Given the CES technology and endogenous agglomeration externalities of the form* $A(S) = e^{\psi S}, \psi > 0$*, cities of different size exist, provided* $\psi > \frac{2(1-\gamma(1-\alpha))}{\overline{M}(1-\alpha)}$.

Proof. See below. ■

This result indicates that agglomeration externalities in production alone can generate the coexistence of cities of different size and productivity. The qualifier requires that for a given size of the labor force $\overline{M}$, the externality must be strong enough. If ψ is high enough, the function $A = e^{\psi S}$ will be convex enough and as a result, there will be a large enough agglomeration effect that generates the existence of multiple equilibria.

Interestingly, a commonly assumed functional form in partial equilibrium, $A(S) = S^{\phi}$, does not generate heterogeneous cities in conjunction with the CES technology. This is true even if A is convex ($\phi > 1$) as shown in the following Corollary. The reason is that already under CES, there is proportionality in the equilibrium demand for labor (proportional across skills), and an externality of this form lifts each city's productivity, but again proportionally. As a result homotheticity, the size of the city is fully governed by the decreasing returns at each skill level. The returns to scale can never be sufficiently strong.

Corollary A. 2 *Given the CES technology and endogenous agglomeration externalities of the form* $A(S) = S^{\phi}, \phi > 0$*, generically cities are of identical size.*

Proof. See below. ■

Ideally we would like to solve the model and prove that multiple equilibria exist also in the presence of extreme-skill complementarities. Unfortunately, that problem is quite a bit more challenging due to the dimensionality of the skill distribution. Not surprisingly, under CES, the proportionality of labor demand implies that distributions are identical across cities. As a result, we only need to solve for the endogenous city size, and not each skill level individually. While we cannot prove any general results, we do conjecture that the nature of the results extends to the non-CES case.

Proofs Agglomeration Externalities

Going back to the system of five equations in the preliminaries, we can now substitute A_j for $A(S_j)$. Denote by $\overline{M} = S_1 + S_2 = \sum_i M_i$ as the economy wide population. Then we will write $S_2 = \overline{M} - S_1$.

From dividing the first by the second and rearranging, we obtain:

$$\left\{ \begin{array}{c} \left[1-\lambda\gamma\left(1-\alpha\right)\right]\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}} \\ -\frac{N_2}{N_1}\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\left[1-\lambda\gamma\left(1-\alpha\right)\right] \end{array} \right\} \left(\frac{M_3}{N_2+N_1\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{1}{\lambda\gamma-1}}}\right)^{\lambda\gamma}\left[y_3+\left(\frac{M_1}{M_3}\right)^{\gamma}y_1\right]^{\lambda}$$

$$= \left\{ \begin{array}{c} \frac{N_2}{N_1}\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\left[1-\gamma\left(1-\alpha\right)\right] \\ -\left[1-\gamma\left(1-\alpha\right)\right]\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}} \end{array} \right\} \left(\frac{M_2}{N_2+N_1\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{1}{\gamma-1}}}\right)^{\gamma} y_2,$$

and from dividing the third by the fourth we have:

$$\left\{ \begin{array}{c} \left(1-\lambda\gamma\right)\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}} \\ -\left(1-\lambda\gamma\right)\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)} \end{array} \right\} \left(\frac{M_3}{N_2+N_1\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{1}{\lambda\gamma-1}}}\right)^{\lambda\gamma}\left[y_3+\left(\frac{M_1}{M_3}\right)^{\gamma}y_1\right]^{\lambda}$$

$$= \left\{ \begin{array}{c} \left(1-\gamma\right)\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)} \\ -\left(1-\gamma\right)\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}} \end{array} \right\} \left(\frac{M_2}{N_2+N_1\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{1}{\gamma-1}}}\right)^{\gamma} y_2$$

Jointly, these two equations give us

$$\frac{\left[1-\lambda\gamma\left(1-\alpha\right)\right]}{\left(1-\lambda\gamma\right)}\frac{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}}-\frac{N_2}{N_1}\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}}{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}}-\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}} = \frac{\left[1-\gamma\left(1-\alpha\right)\right]}{\left(1-\gamma\right)}\frac{\left\{\frac{N_2}{N_1}\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}-\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}\right\}}{\left\{\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}-\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}\right\}} \quad (\bigstar) \tag{A.90}$$

We can now establish the following preliminary results. First, cities are not equal in the number of firms N_j.

Lemma A. 2 *If $\lambda > 1$ and $\lambda\gamma < 1$, then $\frac{N_2}{N_1} \neq 1$, and $Z \neq 1$.*

Proof. Assume $\frac{N_2}{N_1} = 1$, then from ($\bigstar$) we have:

$$\frac{\left[1-\lambda\gamma\left(1-\alpha\right)\right]}{\left(1-\lambda\gamma\right)} = \frac{\left[1-\gamma\left(1-\alpha\right)\right]}{\left(1-\gamma\right)} \tag{A.91}$$

which is a contradiction, since $\lambda > 1 \Rightarrow \frac{[1-\lambda\gamma(1-\alpha)]}{(1-\lambda\gamma)} > \frac{[1-\gamma(1-\alpha)]}{(1-\gamma)}$.

Rewritting the equality (★) above, we obtain:

$$\frac{[1-\lambda\gamma(1-\alpha)]}{(1-\lambda\gamma)}\frac{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}}-\frac{N_2}{N_1}\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}}{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}-\frac{N_2}{N_1}\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}}=\frac{[1-\gamma(1-\alpha)]}{(1-\gamma)}\frac{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}}-\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}}{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}-\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}} \tag{A.92}$$

Then, if $Z=1$, we have $\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}}=\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}=1$. Therefore, we have:

$$\begin{aligned}\frac{[1-\lambda\gamma(1-\alpha)]}{(1-\lambda\gamma)}\frac{\left\{1-\frac{N_2}{N_1}\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}}{\left\{1-\frac{N_2}{N_1}\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}} &= \frac{[1-\gamma(1-\alpha)]}{(1-\gamma)}\frac{\left\{1-\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}}{\left\{1-\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}}\\ \frac{[1-\lambda\gamma(1-\alpha)]}{(1-\lambda\gamma)} &= \frac{[1-\gamma(1-\alpha)]}{(1-\gamma)}\end{aligned}$$

which as we saw before, it is a contradiction. ■

Given Z, we can now establish the main relations between the number of firms N_j, city size S_j, housing prices p_j and TFP $A(S_j)$.

Lemma A. 3 *If $Z<1$, then:*

1. $N_1>N_2$*;*
2. $S_1>S_2$*;*
3. $A_1>A_2$*;*
4. $p_1>p_2$*.*

With opposite inequalities if $Z>1$.

Proof. We establish each of the items in turn

1. Rearranging equality (★), we get:

$$\frac{[1-\lambda\gamma(1-\alpha)]}{(1-\lambda\gamma)}\frac{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}}-\frac{N_2}{N_1}\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}}{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}}-\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}}=\frac{[1-\gamma(1-\alpha)]}{(1-\gamma)}\frac{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}-\frac{N_2}{N_1}\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}}{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}-\frac{p_1}{p_2}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right\}} \tag{A.93}$$

and further simplifying:

$$\frac{[1-\lambda\gamma(1-\alpha)]}{(1-\lambda\gamma)}\frac{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}}\frac{p_2}{p_1}\frac{A(S_1)}{A\left(\overline{M}-S_1\right)}-\frac{N_2}{N_1}\right\}}{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}\frac{p_2}{p_1}\frac{A(S_1)}{A\left(\overline{M}-S_1\right)}-\frac{N_2}{N_1}\right\}}=\frac{[1-\gamma(1-\alpha)]}{(1-\gamma)}\frac{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}}\frac{p_2}{p_1}\frac{A(S_1)}{A\left(\overline{M}-S_1\right)}-1\right\}}{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}\frac{p_2}{p_1}\frac{A(S_1)}{A\left(\overline{M}-S_1\right)}-1\right\}} \tag{A.94}$$

Then:

$$\frac{d}{d\frac{N_2}{N_1}}(LHS)=\frac{[1-\lambda\gamma(1-\alpha)]}{(1-\lambda\gamma)}\frac{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\lambda\gamma}{\lambda\gamma-1}}-\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}\right\}\frac{p_2}{p_1}\frac{A(S_1)}{A\left(\overline{M}-S_1\right)}}{\left\{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A\left(\overline{M}-S_1\right)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}\frac{p_2}{p_1}\frac{A(S_1)}{A\left(\overline{M}-S_1\right)}-\frac{N_2}{N_1}\right\}^2} \tag{A.95}$$

If $Z<1$, we have: $\frac{d}{d\frac{N_2}{N_1}}(LHS)>0$. Then, since $\frac{[1-\lambda\gamma(1-\alpha)]}{(1-\lambda\gamma)}>\frac{[1-\gamma(1-\alpha)]}{(1-\gamma)}$, if $Z<1$ we must have $\frac{N_2}{N_1}<1$. Similarly, if $Z>1$, we have: $\frac{d}{d\frac{N_2}{N_1}}(LHS)<0$. Then, since $\frac{[1-\lambda\gamma(1-\alpha)]}{(1-\lambda\gamma)}>\frac{[1-\gamma(1-\alpha)]}{(1-\gamma)}$, if $Z>1$ we must have $\frac{N_2}{N_1}>1$.

2. From the fifth equation, we have:

$$S_1=(M_1+M_3)\frac{Z^{\frac{1}{\lambda\gamma-1}}}{\frac{N_2}{N_1}+Z^{\frac{1}{\lambda\gamma-1}}}+\frac{Z^{\frac{1}{\gamma-1}}M_2}{\frac{N_2}{N_1}+Z^{\frac{1}{\gamma-1}}} \tag{A.96}$$

If $Z<1$, from the previous Lemma we know that $\frac{N_2}{N_1}<1$. Since:

$$\begin{aligned}\frac{d}{d\lambda}\left(\frac{Z^{\frac{1}{\lambda\gamma-1}}}{\frac{N_2}{N_1}+Z^{\frac{1}{\lambda\gamma-1}}}\right) &= \frac{-\frac{\gamma}{(\lambda\gamma-1)^2}Z^{\frac{1}{\lambda\gamma-1}}\ln Z\times\left(\frac{N_2}{N_1}+Z^{\frac{1}{\lambda\gamma-1}}\right)+\frac{\gamma}{(\lambda\gamma-1)^2}Z^{\frac{1}{\lambda\gamma-1}}Z^{\frac{1}{\lambda\gamma-1}}\ln Z}{\left(\frac{N_2}{N_1}+Z^{\frac{1}{\lambda\gamma-1}}\right)^2}\\ &= -\frac{\frac{\gamma}{(\lambda\gamma-1)^2}\frac{N_2}{N_1}Z^{\frac{1}{\lambda\gamma-1}}\ln Z}{\left(\frac{N_2}{N_1}+Z^{\frac{1}{\lambda\gamma-1}}\right)^2}>0 \text{ since } \ln Z<0 \text{ as } Z<1\end{aligned}$$

We have:

$$S_1>(M_1+M_3+M_2)\frac{Z^{\frac{1}{\gamma-1}}}{\frac{N_2}{N_1}+Z^{\frac{1}{\gamma-1}}}>\frac{\overline{M}}{2} \tag{A.97}$$

The same logic establishes the opposite when $Z>1$.

3. From the previous lemma, if $Z<1$, we have that $S_1>S_2$. Since $A'(\cdot)>0$, $A(S_1)>A\left(\overline{M}-S_1\right)=A_2$.

4. But then, from (★):

$$\frac{N_1}{N_2}\frac{A(S_1)}{A\left(\overline{M}-S_1\right)}\frac{\left\{\begin{array}{c}[1-\lambda\gamma(1-\alpha)]Z^{\frac{\lambda\gamma}{\lambda\gamma-1}}\left(\frac{M_3}{N_2+N_1Z^{\frac{1}{\lambda\gamma-1}}}\right)^{\lambda\gamma}\left[y_3+\left(\frac{M_1}{M_3}\right)^{\gamma}y_1\right]^{\lambda}\\ +[1-\gamma(1-\alpha)]Z^{\frac{\gamma}{\gamma-1}}\left(\frac{Z^{\frac{1}{\gamma-1}}M_2}{N_2+N_1Z^{\frac{1}{\gamma-1}}}\right)^{\gamma}y_2\end{array}\right\}}{\left\{\begin{array}{c}[1-\lambda\gamma(1-\alpha)]\left(\frac{M_3}{N_2+N_1Z^{\frac{1}{\lambda\gamma-1}}}\right)^{\lambda\gamma}\left(y_3+\left(\frac{M_1}{M_3}\right)^{\gamma}y_1\right)^{\lambda}\\ +[1-\gamma(1-\alpha)]\left(\frac{M_2}{N_2+N_1Z^{\frac{1}{\gamma-1}}}\right)^{\gamma}y_2\end{array}\right\}}=\frac{p_1}{p_2} \tag{A.98}$$

Since $Z<1$, we showed that $\frac{N_1}{N_2}>1$. Then:

$$\frac{p_1}{p_2} > \frac{N_1}{N_2}\frac{A\left(S_1\right)}{A\left(\overline{M}-S_1\right)}Z^{\frac{\gamma}{\gamma-1}}. \tag{A.99}$$

Similarly for $Z > 1$.

■

Next, we establish the result for thicker tails for the case of endogenous TFP from agglomeration externalities.

Lemma A. 4 *Given $\lambda > 1$ and $\lambda\gamma < 1$, the larger city has thicker tails.*

Proof. If $Z < 1$, from Lemma A.3 we know that city 1 is larger than city 2, and that $A_1 > A_2$. Therefore we can apply Theorem 2: city 1 is larger an has thicker tails. Instead, if $Z > 1$, we know that city 2 is larger than city 1, and that $A_1 < A_2$. Now we can define $Z' = 1/Z$ (or relabel the cities) and again apply Theorem 2: city 2 is larger and has thicker tails. ■

Now our main result immediately follows from Lemmas A.2, A.3, and A.4:

Theorem A.3 *Given endogenous agglomeration externalities and given $\lambda > 1$ and $\lambda\gamma < 1$, and provided cities of different size exist, the larger city has:*

We now establish the proof of Theorem A.4

Theorem A.4 *Given the CES technology and endogenous agglomeration externalities of the form $A(S) = e^{\psi S}, \psi > 0$, cities of different size exist, provided $\psi > \frac{2(1-\gamma(1-\alpha))}{\overline{M}(1-\alpha)}$.*

Proof. In the case of CES ($\lambda = 1$) we can write the system of 9 equilibrium equations as:

$$\begin{cases} m_{12} = \frac{M_1}{N_1\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A(S_2)}{A(S_1)}\right]^{\frac{1}{\gamma-1}}+N_2} \\ m_{22} = \frac{M_2}{N_1\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A(S_2)}{A(S_1)}\right]^{\frac{1}{\gamma-1}}+N_2} \\ m_{32} = \frac{M_3}{N_1\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A(S_2)}{A(S_1)}\right]^{\frac{1}{\gamma-1}}+N_2} \\ \left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A(S_2)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}\left\{\left(m_{12}\right)^{\gamma}y_1+\left(m_{32}\right)^{\gamma}y_3+\left(m_{22}\right)^{\gamma}y_2\right\} = \left[\frac{H}{N_1}-k\right]\frac{p_1}{\alpha\gamma A(S_1)} \\ \left\{\left(m_{12}\right)^{\gamma}y_1+\left(m_{32}\right)^{\gamma}y_3+\left(m_{22}\right)^{\gamma}y_2\right\} = \left[\frac{H}{N_2}-k\right]\frac{p_2}{\alpha\gamma A(S_2)} \\ \left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A(S_2)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}\left\{\left(m_{12}\right)^{\gamma}y_1+\left(m_{32}\right)^{\gamma}y_3+\left(m_{22}\right)^{\gamma}y_2\right\} = \frac{k}{(1-\gamma)A(S_1)}p_1 \\ \left\{\left(m_{12}\right)^{\gamma}y_1+\left(m_{32}\right)^{\gamma}y_3+\left(m_{22}\right)^{\gamma}y_2\right\} = \frac{k}{(1-\gamma)A(S_2)}p_2 \\ S_1 = \left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A(S_2)}{A(S_1)}\right]^{\frac{1}{\gamma-1}}\left[m_{12}+m_{22}+m_{32}\right]N_1 \\ S_2 = \left[m_{12}+m_{22}+m_{32}\right]N_2 \end{cases} \tag{A.100}$$

From eqs. (6) and (7), we obtain:

$$\frac{\left[\left(\frac{p_1}{p_2}\right)^{\alpha}\frac{A(S_2)}{A(S_1)}\right]^{\frac{\gamma}{\gamma-1}}\left\{\left(m_{12}\right)^{\gamma}y_1+\left(m_{32}\right)^{\gamma}y_3+\left(m_{22}\right)^{\gamma}y_2\right\}}{\left\{\left(m_{12}\right)^{\gamma}y_1+\left(m_{32}\right)^{\gamma}y_3+\left(m_{22}\right)^{\gamma}y_2\right\}} = \frac{kp_1}{\left(1-\gamma\right)A\left(S_1\right)}\times\frac{\left(1-\gamma\right)A\left(S_2\right)}{kp_2}, \tag{A.101}$$

and after rearranging:

$$\frac{p_1}{p_2} = \left(\frac{A(S_1)}{A(S_2)}\right)^{\frac{1}{1-\gamma(1-\alpha)}} \tag{A.102}$$

From (4) and (6), and from (5) and (7) we have:

$$N_1 = \frac{H}{\left[1 + \frac{\alpha\gamma}{(1-\gamma)}\right] k} = N_2 \tag{A.103}$$

Substituting $(1), (2)$, and (3) into (8), and using the price ratio $\frac{p_1}{p_2}$ we get:

$$S_1 = \left\{ \frac{\left(\frac{A(S_1)}{A(S_2)}\right)^{\frac{(1-\alpha)}{1-\gamma(1-\alpha)}}}{1 + \left(\frac{A(S_1)}{A(S_2)}\right)^{\frac{(1-\alpha)}{1-\gamma(1-\alpha)}}} \right\} \overline{M} \quad \text{and} \quad S_2 = \left\{ \frac{1}{1 + \left(\frac{A(S_1)}{A(S_2)}\right)^{\frac{(1-\alpha)}{1-\gamma(1-\alpha)}}} \right\} \overline{M} \tag{A.104}$$

Then:

$$\frac{S_1}{S_2} = \left(\frac{A(S_1)}{A(S_2)}\right)^{\frac{(1-\alpha)}{1-\gamma(1-\alpha)}} \tag{A.105}$$

Since $S_2 = \overline{M} - S_1$, we have:

$$\frac{S_1}{\overline{M} - S_1} = \left(\frac{A(S_1)}{A\left(\overline{M} - S_1\right)}\right)^{\frac{(1-\alpha)}{1-\gamma(1-\alpha)}} \tag{A.106}$$

Now consider the case where $A(S) = e^{\psi S}$ and denote the exponent on the RHS term by $K = \frac{\psi(1-\alpha)}{1-\gamma(1-\alpha)}$ and observe that it is positive. Then the equilibrium condition is:

$$\begin{aligned} \frac{S_1}{\overline{M} - S_1} &= \left(e^{2S_1 - \overline{M}}\right)^K \\ \log\left(\frac{S_1}{\overline{M} - S_1}\right) &= K\left(2S_1 - \overline{M}\right) \end{aligned}$$

First, there is always a symmetric equilibrium $S_1 = \frac{\overline{M}}{2}$. Substituting $S_1 = \frac{\overline{M}}{2}$ gives 0 both on the LHS and the RHS.

Next, we show that there are also two asymmetric equilibria, one where $S_1 > \frac{\overline{M}}{2} > S_2$ and the mirror image with $S_2 > \frac{\overline{M}}{2} > S_1$. To see this, observe that the RHS is linear with bounded support on $[0, \overline{M}]$ and takes values between $-K\overline{M}$ and $K\overline{M}$. The LHS takes values between $-\infty$ and $+\infty$: at $S_1 = 0$, the LHS is equal to $\log 0 = -\infty$ and at $S_1 = \overline{M}$, the LHS is equal to $\log \infty = +\infty$. The slope of the LHS is positive and given by

$$\frac{\overline{M}}{S_1(\overline{M} - S_1)}. \tag{A.107}$$

We know that there is an intersection at $S_1 = \frac{\overline{M}}{2}$, and therefore, given the behavior at $S_1 = 0$ and ∞ and continuity of both LHS and RHS, there is are at least two more intersections provided the slope at $S_1 = \frac{\overline{M}}{2}$ is

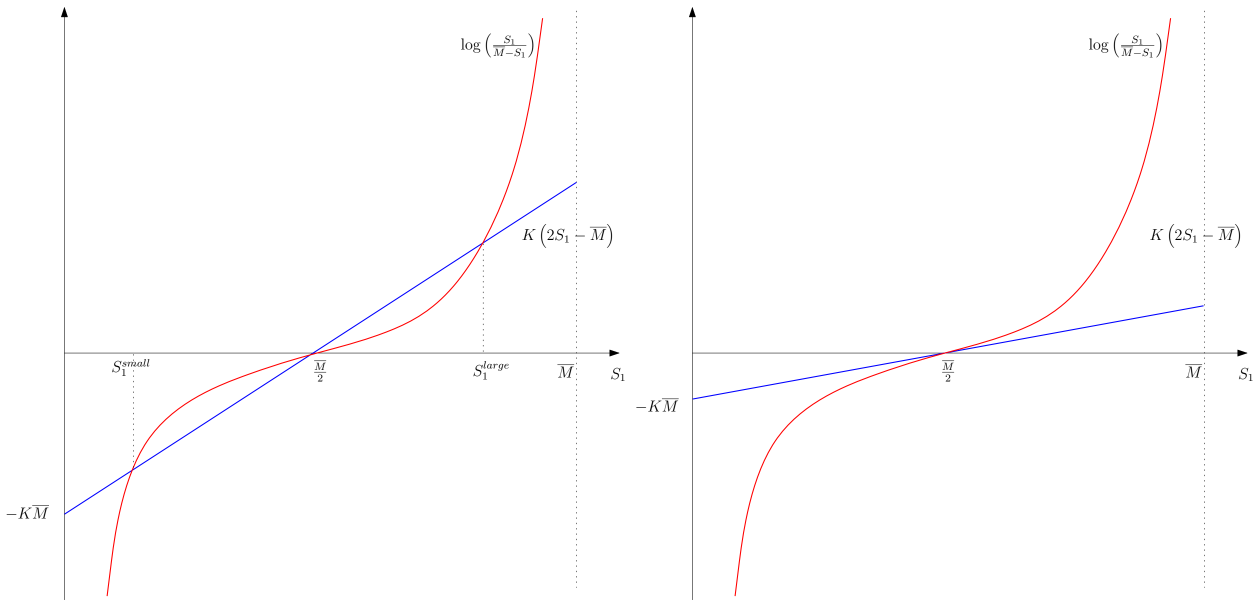

Figure 15: Proof of Theorem A.4: A. Multiple equilibria with cities of different sizes exist when K is large enough; B. a unique equilibrium exists with identical cities exists when K is small.

flatter than the slope of the RHS, i.e. provided:

$$\frac{4}{\overline{M}} < 2K \quad \text{or} \quad \psi > \frac{2\left(1-\gamma\left(1-\alpha\right)\right)}{\overline{M}(1-\alpha)}. \tag{A.108}$$

The logic is illustrated in Figure 15. $\blacksquare$

Corollary A.2 *Given the CES technology and endogenous agglomeration externalities of the form $A(S) = S^{\phi}, \phi > 1$, generically cities are of identical size.*

Proof. Now the equilibrium condition can be written as:

$$\frac{S_1}{\overline{M}-S_1} = \left(\frac{S_1}{\overline{M}-S_1}\right)^{\frac{\phi(1-\alpha)}{1-\gamma(1-\alpha)}}. \tag{A.109}$$

which has a unique solution $S_1 = \overline{M} - S_1$ provided $\frac{\phi(1-\alpha)}{1-\gamma(1-\alpha)} - 1 \neq 0$. When $\frac{\phi(1-\alpha)}{1-\gamma(1-\alpha)} - 1 = 0$, there is indeterminacy in the size of both cities and $S_1 \in [0, \overline{M}]$. However, this configuration of parameters is non-generic, therefore generically $S_1 = \frac{\overline{M}}{2}$ and cities are identical. $\blacksquare$