Inequality Across and Within US Cities around the Turn of the Twenty-First Century

Felipe Schwartzman

Y ou have just finished your PhD, and you have excellent offers. One of them is in a major city such as Washington, DC, and the other is in a smaller place like Charlottesville, Virginia, or Durham, North Carolina. In terms of the quality of each department, all offers look like excellent moves, and your advisor would be thrilled to see you in either place. It comes down to where you would rather live. If you move to the larger city, opportunities might look better down the line. There are multiple great universities in and around Washington, increasing the number of people you can interact with and learn from. International organizations in Washington, such as the IMF and the World Bank, are willing to pay high salaries to people with your qualifications, bringing up your market wage. At the same time, a large city offers unique amenities like superb restaurants and great art. On the other hand, rent is expensive: you will probably need to settle for a smaller house or a longer commute.

Suppose instead that you have never pursued a PhD. In fact, you barely graduated from high school. Your choices may look fairly different. Washington, DC, will have few, if any, good jobs for you, since there is no space for the large industrial facilities that are likely to offer good jobs for people without college degrees. You do not have much interest in the elitist art emphasized in posh neighborhoods and cer-

tainly no disposable income to go to nice restaurants. Smaller towns may offer you better prospects, and you can purchase a better house there with your wage.

In the end, irrespective of your education, all of these factors are also influenced by your own personal preferences. Durham might be closer to relatives, or you might have a special appreciation for monuments and memorials on the National Mall.[1]

Together, all these factors determine a spatial equilibrium, in which people choose where to live and wages and rental prices adjust accordingly. What recent research has shown is that the nature of this spatial equilibrium has changed in the US in the past few decades. Large cities such as Washington, DC, New York, or San Francisco are increasingly places for the skilled elites. Those cities have experienced higher wage growth, and their growth has been unequal; concentrating income among educated professionals. At the same time, much of that wage growth has been offset by increased rents. Not surprisingly, the share of college-educated workers in these cities has increased.

Those facts can be accounted for by a spatial equilibrium framework as consequences from relative increases in the demand for skilled labor by firms in large, skilled cities. An important part of the trends may have to do with adoption of computer technology. The mechanism is also most likely related to greater spillovers among those skilled workers in those cities, but the precise nature of those spillovers is still open to more research.

In what follows, I describe in greater detail the research documenting those facts and the lessons that one can derive about the underlying mechanisms. In Section 1, I lay out the facts, and in Section 2, I lay out the explanations. Section 2 includes the presentation of a canonical urban equilibrium model with two occupations that can be used to think through different mechanisms and discuss the evidence surrounding alternative hypotheses.

1.   KEY FACTS

The key facts about inequality across and within US cities can be summarized as follows: if you have a college degree or, more generally, are a more skilled worker, you are more likely to live in a larger city. Wages in those cities are generally higher, although they are also offset by higher rental prices. Those relative wage gains are particularly pronounced

---

[1] Perhaps the answer is to go for the happy medium and come work in Richmond, Virginia!

among more highly skilled workers, making those large, skilled cities more unequal.

The literature typically identifies skills with individual characteristics correlated with productivity. High-skilled workers can be identified as ones who have high levels of education, work in high-wage occupations such as management or law, or work in high-wage industries such as professional and business services. One alternative to this vertical definition of skill is adopted by Bacolod, Blum, and Strange (2009), who examine how salaries vary with different skill dimensions such as cognitive processing or personal interaction. Alternatively, skills can be estimated from structural models (Baum-Snow and Pavan 2013; Eeckhout, Pinheiro, and Schmidheiny 2014).

In spite of this variety of measures, different papers have recently documented fairly robust facts about spatial inequality in the US around the turn of the twenty-first century:

1. Larger cities have a greater concentration of high-skilled workers.

This fact is true regardless of how one measures skill. Baum-Snow and Pavan (2013), Eeckhout et al. (2014), and Davis and Dingel (2017) provide evidence for measures of skill quality based on education and occupation. For example, Baum-Snow and Pavan (2013) report that in the late 2000s, about 40 percent of the workforce of cities in the top size decile had a college degree, whereas in cities in the bottom size decile only about 20 percent of the workforce had a college degree.

The extent to which these relationships have changed since the 1980s appears to be dependent on the exact definition of skills. Berry and Glaeser (2005) and Diamond (2016) find that cities with a high share of college graduates have experienced a larger increase in that share, but Baum-Snow, Freedman, and Pavan (forthcoming) argue that such a relationship is not apparent if skilled workers are redefined to include those with "some college." This seems to suggest that much of the change over this period has occurred due to more people finishing college. The sensitivity of changes over time to definitions is probably also a reflection of high persistence of the educational composition of cities, a fact emphasized by Beaudry, Doms, and Lewis (2010).

Such a sensitivity of time trends to the boundaries between high- and low-skilled workers appears to call for a more disaggregated view. Disaggregation reveals that in recent periods both the highest- and lowest-skilled workers tend to concentrate in large cities, with smaller cities exhibiting a more concentrated skill distribution. For example, when measuring skills by education, Eeckhout et al. (2014) find that large cities include more college graduates, but also more high school dropouts. Of the latter, many (but not all) appear to be recent

international immigrants.[2] The differences in dispersion were not always present. Davis and Dingel (2017) show evidence to the effect that this phenomenon is relatively new, with larger cities exhibiting uniformly more skilled workers in 1980. This move toward a more extreme distribution of skills in recent decades is consistent with Autor and Dorn's (2013) description of a geographic dimension to job market polarization, with cities that have a high share of workers in occupations with intermediate wage levels ("routine intensive" occupations) seeing large shifts in their labor composition toward low-wage occupations ("service" occupations).

2. Nominal wages are overall higher and increasing in larger and in more skill-intensive cities, but real wages are not necessarily.

A key distinction when interpreting geographic data is between nominal and real wages (or income, more broadly), where the latter incorporates a local price adjustment. Because not all goods consumed by households can be freely traded between cities, the law of one price does not necessarily hold everywhere. In particular, land is the ultimate nontradable good and corresponds to a large fraction of households' consumption baskets.

One of the most important stylized facts of urban economics is that larger cities exhibit higher nominal wages. Most recently, Baum-Snow et al. (forthcoming) have calculated that from 2005–07 nominal wages increase 0.065 percent for each percentage increase in city size. Relatedly, Glaeser and Maré (2001) pin the wage differences between urban and rural areas to around 33 percent in 1990. Those relationships have strengthened over time. The elasticity of wages to city size reported by Baum-Snow et al. (forthcoming) for 2005–07 is about 50 percent larger than what they report for 1980. Also, in the working paper version of their 2013 paper, Baum-Snow and Pavan point out that the wage gap between the largest cities (1.5 million people or more) and rural areas increased from 24 percent in 1980 to 33 percent in 2000. These gaps reflect in part the increasing correlation between city size and skill mix, but after controlling for those they remain sizable at 17 percent and 24 percent, respectively, and the trend remains noticeable. There is, moreover, a strengthening of the relationship between the skill intensity of a city and wages, with more skill-intensive cities exhibiting higher wages for both skilled and unskilled workers (Diamond 2016).

---

[2] Interestingly, they do not find any differences in average skills between cities even as they confirm the findings by others that those cities feature a higher concentration of college-educated workers. This is because college-educated workers concentrate in the higher quantiles of the skill distribution.

While the relationship between nominal wages and city size is a clear and robust fact of urban economics, this relationship does not necessarily extend itself to real wages. For a recent example, Eeckhout et al. (2014) report that there is no systematic difference in average real wages between cities. Moretti (2013) and Diamond (2016) show, moreover, that cities with a large share of skilled workers are also cities in which rent prices are higher and have increased the most in recent decades. In effect, Moretti (2013) shows that, while inequality of nominal wages across cities has clearly increased, it was met by an increased dispersion in rents, so cross-city inequality in real wages has not increased as strongly.

One important caveat to those findings is that the measurement of local price levels is itself fraught. In a recent paper, Handbury and Weinstein (2015) show that typical price indices measured to compare standards of living across cities are biased because they do not properly account for differences in the quality and variety of goods. They find that after one properly controls for those, there is a negative relationship between the price of tradable goods and city size. Given existing evidence, this would imply real wages that increase with city size.

3. The skill premium is higher and increasing in larger cities or cities with more skilled workers.

Larger cities appear to be more unequal. When comparing rural areas and the three largest metropolitan areas, Baum-Snow and Pavan (2013) report that in 2004–07, the variance of log hourly wages was 0.28 in the former and 0.53 in the latter. This, they show, is a relatively new phenomenon, since in 1979 the variances of log hourly wages for rural areas and the three largest metropolitan areas were 0.19 and 0.24, respectively.

A major focus of the recent literature has been the evolution of the skill premium across and within cities. One robust finding is that wage premia increase with city size. Eeckhout et al. (2014) and Davis and Dingel (2017) report those relationships for recent data using a variety of skill measures based on education, occupation, or observed real wages. Baum-Snow et al. (forthcoming) also point out that the relationship between skill premium and city size has become more pronounced over time, doubling in its strength over that period.

Finally, recent data also show more skilled cities exhibiting larger skill premia (Moretti 2013), although the fact does not appear to be robust to the exact definition of skill. For example, Beaudry et al. (2010) and Hendricks (2011) do not appear to find a robust relationship in recent data. There appears to be more consensus around an increasingly positive relationship between skill composition and skill premia over time. In fact, Beaudry et al. (2010) show evidence that

skilled cities had lower skill premia in 1980 and before, but this negative correlation disappeared in the early 2000s. Together, these facts point to the aggregate inequality trends as having an important geographic component, with large, skill-intensive cities leading the charge.[3]

Implications for Interpretation of Wage Inequality

The facts above suggest a reinterpretation of observed trends in wage inequality. A high wage in New York will sound appealing until one realizes how much one needs to pay for rent. Since both wages and rents are higher in large, skill-intensive cities (Fact 2), this suggests that adjusting for the local cost of living could imply less inequality in standards of living than is implied by wages alone, a point explored by Moretti (2013). He finds that the real wage differential between college and noncollege workers has increased 20 to 30 percent less than the nominal wage differentials, as rents have increased more quickly in skill-intensive cities.

At the same time, New York may offer more than smaller cities in terms of the quality of its restaurants and art scene. Diamond (2016) presents evidence that the increase in rental prices was more than matched by an increase in the amenities provided to residents of more skill-intensive cities. She estimates the effect based on a structural model similar to one we will present in Section 2 below, but in her model, local amenities change endogenously in response to the population composition. She estimates the model using measures of amenities such as quality of public schools, crime rates, and restaurant density, and she finds that once one accounts for those effects, the inequality of standards of living increases by 30 percent more than what is implied by wage inequality trends alone.

2.   EXPLAINING THE FACTS

The most natural explanation for the set of facts described above, advanced by Berry and Glaeser (2005), is that the demand for skilled workers has increased more in cities that are larger and more skill intensive, while the demand for unskilled workers has not increased much

---

[3] Again, the direction and strength of correlations appears to depend on exactly how skill is measured. Moretti (2013) finds a very clear positive correlation between college share and college premium in 2000, whereas Beaudry et al. (2010) and Hendricks (2011) do not find a statistically significant positive correlation between a "college equivalent" share and returns to education. The main difference appears to be again in how skilled labor is defined.

anywhere. In this section, we explain why such a view is a natural fit for the data. We will then examine different theories behind that increase in demand, including the rise of computers and externalities.

The most important alternative to labor demand increases is an increase in endogenous sorting for unobserved worker characteristics, so that, for example, among college-educated workers, it is the most productive ones who choose to live in large cities such as New York and San Francisco. Nevertheless, the most recent literature appears to indicate that such sorting is unlikely to be an important driving force behind the observed facts.

In order to build the argument, we rely on a class of equilibrium models that have been originally proposed by Rosen (1979) and Roback (1982). In those models, cities exist in fixed locations and are characterized by a production technology for a fully tradable good and by their land availability. At a given wage, firms in more productive locations seek to attract more workers. However, as workers move into those cities, their demand for housing pushes rents up. Because workers are free to choose where to live, they will only choose to live in cities with high rents if wages are commensurably high. In spatial equilibrium, rents in more productive cities are just high enough to offset the productivity advantage of firms in those cities. The congestion coming from scarce land ensures that all cities are populated in equilibrium, irrespective of the productivity of their workers.

We now build a variant of such a model with workers of different skills. Similar variants have been used in recent work by Moretti (2013) and Diamond (2016), among others.

Model Setup

There are N cities, indexed $n \in \{1, ..., N\}$. Each of these cities is equipped with a production technology for a tradable good that depends on the number of high- and low-skilled workers in the city. A representative firm in city n can produce quantity Y_n of the final good by employing L_n^H high-skilled workers and L_n^L low-skilled workers according to the constant returns to scale production function:

$$Y_n = F_n(L_n^L, L_n^H).$$

Note that we allow the production function to be city-specific. Differences in the production function may also lead certain cities to produce the final good using more of one or the other type of labor. Those differences can capture "natural advantages" that can make a city more productive than another. The clearest examples of such advantages include proximity to waterways or to fertile terrain, but one could be

naturally skeptical as to whether such natural advantages directly explain the productivity differences between modern cities. In subsequent discussions on the causal mechanisms behind the facts surrounding inequality and geography, we allow for variation in capital stock and for externalities.

Also note that we assume firms are native to individual cities and stay there. This goes counter to a long-standing emphasis of the spatial economics literature on location decisions of firms. The assumption is inoccuous, however, because of our assumptions of constant returns to scale and of city-specific production technology. One could similarly postulate a model where individual firms are free to establish themselves in any city and produce using the local technology. In equilibrium, the zero-profit condition would imply the same spatial distribution of production. A less inoccuous alternative, which we do not explore, would be to allow entrepreneurs with different abilities to choose which city to live in. For an example of a framework with this feature, see Behrens, Duranton, and Robert-Nicoud (2014).

Labor markets are competitive, so firms pay wages equal to the marginal product of labor. For each skill level $k \in \{L, H\}$, this induces the labor demand equation:

$$w_n^k = \frac{\partial F_n(L_n^L, L_n^H)}{\partial L_n^k},$$

where w_n^k is the wage of workers of type k in city n in terms of the final tradable good.

Workers have preferences over the tradable good, housing, and location. For a given worker i, those idiosyncratic experiences are captured by vector worker-specific amenity parameters $\{\varepsilon_1(i), \varepsilon_2(i), ..., \varepsilon_N(i)\}$, where $\varepsilon_1(i)$ parameterizes the worker-specific preference for living in city n. Those capture the extent to which different workers have preferences for different cities. They can incorporate, for example, proximity to family or to the place where the worker grew up. A worker indexed i with skill $k \in \{L, H\}$ living in city n enjoys a utility equal to:

$$U_n^k(i) = G_n(X_n^k, H_n^k, \varepsilon_n(i)),$$

where X_n^k and H_n^k are, respectively, the amount a worker in city n consumes of the tradable good and of housing. Note again that the worker's utility function is allowed to depend on the city n where the household chooses to live. This captures the notion that amenities, such as weather, can make some cities overall more pleasant than others. Those amenity effects operate in addition to the idiosyncratic preference shifts captured by $\varepsilon_n(i)$. Understanding the origin of such

amenities and the extent to which they are endogenous has also been an important topic of research (see Albouy [2012] and Diamond [2016] for recent contributions).

As we will see, preferences for housing play an important role in the model: they introduce a source of congestion at the city level, generating a reason for population to spread across cities in spite of differences in the marginal product of workers. Similar sources of congestion could arise in the presence of other nontradable goods (such as certain traditional services) or trade costs generating home bias in consumption. Other "nonpecuniary" sources of congestion, including pollution and crime, could be captured by endogenizing the amenity parameters.

The workers' purchases of final goods and housing has to satisfy their budget constraints. We assume the only income workers receive is their wage, so that

$$X_n^k + r_n H_n^k \leq w_n^k,$$

where r_n is the rental price of housing, quoted relative to the tradable good. We assume workers can rent houses but not buy them. In a static framework such as the one presented here, this difference is mostly unconsequential. In a dynamic framework, the difference matters since the wealth of workers who purchase housing would become a function of the history of shocks to housing values in the cities where they lived. This has the potential to generate interesting effects over the wealth distribution but would be computationally challenging and has yet to be extensively explored in the literature.

Workers can freely choose in which city to live. We can solve their problem in two stages: first, we solve for the optimal choice of housing and final goods consumption given that the household lives in some city n. This induces a value function $V_n^k(r_n, w_n^k, \varepsilon_n(i))$, satisfying:

$$V_n(r_n, w_n^k, \varepsilon_n(i)) = \max G_n(X_n^k, H_n^k, \varepsilon(i)) \text{ s.t. } X_n + r_n H_n \leq w_n^k.$$

Given those value functions, the household then selects as its living location the city where it attains the highest value. This induces a labor supply as a function of wages and rental prices. The number of workers of type k in city n is thus given by the fraction of workers who have a draw of the worker-specific amenity parameter $\{\varepsilon_1(i), \varepsilon_2(i), ..., \varepsilon_N(i)\}$, such that

$$L_n^k = \Pr \left[V_n(r_n, w_n^k, \varepsilon_n(i)) > \max_{n' \neq n} V_n(r_{n'}, w_{n'}^k, \varepsilon_{n'}(i)) \right] \bar{L}^k,$$

where $\bar{L}^k$ is the total number of workers of type k.

The model is closed by housing market clearing conditions, implying that housing demand within each city has to be equal to housing supply within each city.

$$L_n^H H_n^H + L_n^L H_n^L = \bar{H}_n, \tag{1}$$

where $\bar{H}_n$ is the supply of housing in city n. Finally, we assume that all rental income is appropriated by absentee landlords who only have preferences for the final good and do not supply labor. This assumption ensures that the market for the tradable good clears.

Model Parameterization

In what follows we argue that Facts 1, 2, and 3 in Section 1 can be largely explained by cross-city variations in the demand for skilled labor. In particular, we specialize the model by assuming that all differences in production functions across locations stem from a skilled-labor augmenting component:

$$F_n\left(L_n^L, L_n^H\right) = \left[\left(L_n^L\right)^{\frac{\theta-1}{\theta}} + \left(\lambda_n L_n^H\right)^{\frac{\theta-1}{\theta}}\right]^{\frac{\theta}{\theta-1}}, \ \theta > 1, \tag{2}$$

where θ is the elasticity of substitution between different skill levels, and λ_n is a labor-augmenting parameter specific to the labor of college-educated workers. The assumption $\theta > 1$ is consistent with common estimates of the elasticity of substitution between different types of labor.[4] Under this parameterization, labor demand functions become:

$$L_n^L = \left(w_n^L\right)^{-\theta} Y_n \tag{3}$$

$$L_n^H = \left(\lambda_n\right)^{\theta-1} \left(w_n^H\right)^{-\theta} Y_n. \tag{4}$$

We parameterize the household utility of different types of workers k living in different cities n as:

$$G_n^k(X_n^k, H_n^k, \varepsilon_n(i)) = A_n \varepsilon_n(i) \left(\frac{X_n^k}{1-\beta}\right)^{1-\beta} \left(\frac{H_n^k}{\beta}\right)^{\beta}.$$

The parameterization assumes a unit elasticity of substitution between final goods and housing. The term $A_n \varepsilon_n(i)$ captures differences in amenities between cities, with A_n incorporating amenities that affect

---

[4] See Ciccone and Peri (2005) for a useful summary of estimates of the elasticity of substitution between college- and noncollege-educated workers.

all workers (such as climate or quality of public schools) and $\varepsilon_n(i)$ incorporating worker-specific amenities (such as proximity to family).

Under this parameterization, the value function for a household living in city n can be written as:

$$V_n^k(r_n, w_n^k, \varepsilon_n(i)) = \varepsilon_n(i) A_n \frac{w_n^k}{(r_n)^\beta}.$$

We also assume that $\varepsilon_n(i)$ has a Fréchet distribution with shape parameter ν and is drawn independently for each city. This distribution is commonly used in trade models and has the property that it is stable under the max operator, i.e., the max of two random variables that have a Fréchet distribution is also distributed according to a Fréchet. This property makes it particularly convenient for use in aggregate models where agents make discrete choices. Furthermore, under certain conditions, it emerges naturally as the limiting distribution for the max of a sequence of random variables. It can therefore be motivated by the notion that, when considering a given city, individuals are also choosing the best of several living situations that they have available to them within that city.

As we show in the Appendix, one can then derive the labor supply function as:

$$L_n^k = \frac{\left(A_n \frac{w_n^k}{(r_n)^\beta}\right)^\nu}{\sum_{n'} \left(A_{n'} \frac{w_{n'}^k}{(r_{n'})^\beta}\right)^\nu} \bar{L}^k. \tag{5}$$

The parameter ν controls the degree of heterogeneity in tastes. This in turn governs the supply elasticity of the labor supply. A high value of ν corresponds to low heterogeneity. Thus, small variations in wages received (or rents paid) by workers in some city n imply large changes in the number of workers willing to live in that city. In particular, for the extreme case in which there is no heterogeneity in tastes, $(\nu \to \infty)$, all workers have to be indifferent between all locations, so $A_n w_n^k / r_n^\beta$ is the same for all n. Conversely, a low value of ν corresponds to high heterogeneity. In that case, most workers choose where to live based entirely off their idiosyncratic preferences, and variations in the wages (or rents paid) have little bearing on the number of workers living in each city.

In the Appendix, we also show that the average utility of a household that chooses to live in city n is

$$V^k = \Gamma \left(1 - \frac{1}{\nu} \right) \left[\sum_n \left(A_n \frac{w_n^k}{(r_n)^\beta} \right)^\nu \right]^{\frac{1}{\nu}},$$

where Γ is a gamma function. Note that the average utility does not depend on the location, so that we can denote average worker welfare V^k without the city subscript. The reason is that as real wages increase in a city, workers who have lower amenity values for that city decide to live there, so that high real wage cities will include more workers with low idiosyncratic preferences for that city. Given the Fréchet distribution of tastes, the positive impact of the real wage on city welfare is exactly offset by the negative impact of worker selection.

This calculation allows us to write the labor supply condition more compactly as:

$$L_n^k = \left(\Gamma \left(1 - \frac{1}{\nu} \right) \frac{w_n^k}{V^k (r_n)^\beta} \right)^\nu \bar{L}^k. \tag{6}$$

Finally, we assume that housing supply does not vary across cities, $\bar{H}_n = 1$. This rules out land endowment as a key determinant of city size in the model and is consistent with the casual observation that some of the largest cities in the US, such as New York or San Francisco, are confined on relatively small land masses.[5]

Explaining the Facts with Variation in Skilled-Labor Augmenting Technology

We now show how, at least qualitatively, one can explain the facts in Section 1 entirely as a function of variations in demand. In order to do this, we make the stark assumption that $A_n = 1$ for all n, so that cities do not differ by intrinsic amenities. The only exogenous difference between cities is thus given in the skilled-labor augmenting productivity parameter λ_n. That serves as a shifter in the demand for skilled labor in different cities.

Algebraically, the way in which variation in λ_n implies Facts 1 and 2 is easiest to see in the case of homogeneous preferences ($\nu \to \infty$).[6] In that case, inspection of equation (6) implies that real wages have to

---

[5] Hsieh and Moretti (2017) explore the effect of land restriction regulations on city size and find that because of those restrictions, there is less concentration of population in large cities than there should be, leading to substantial output losses.

[6] See the Appendix for a derivation of the results in the general heterogeneous case.

be the same in all locations; that is, for all n where workers choose to live, it must be the case that, for $k \in \{L, H\}$ and $n \in \{1, ..., N\}$,

$$\frac{w_n^k}{r_n^\beta} = V^k. \tag{7}$$

Substituting this expression in the labor demand equations (3) and (4), we have that:

$$L_n^L = \left(V^L r_n^\beta\right)^{-\theta} Y_n,$$

$$L_n^H = (\lambda_n)^{\theta-1} \left(V^H r_n^\beta\right)^{-\theta} Y_n.$$

Taking the ratio of both labor demand functions, we have that $L_n^H / L_n^L \propto (\lambda_n)^{\theta-1}$, so that with $\theta > 1$, cities in which skilled workers are more productive will have a greater proportion of those workers. Substituting the labor demand equations into the CES production function 2, canceling out Y_n, and rearranging yields an expression for rents in each city n,

$$r_n = \left[\left(V^L\right)^{1-\theta} + (\lambda_n)^{\theta-1} \left(V^H\right)^{1-\theta}\right]^{\frac{1}{\beta}\frac{1}{\theta-1}}.$$

It follows that, along the cross-section of cities, rents increase in λ_n. From the indifference condition (7), nominal wages for both worker types must increase with rents. Thus, cities with a higher fraction of skilled workers are also cities with higher nominal wages for all workers, consistent with Fact 2. At the same time, real wages in those cities are not necessarily higher.

Optimal worker demand for housing implies they will spend a fraction β of their income on housing. Using the indifference condition (7), we can write housing demand in terms of rents only:

$$H_n^k = \beta V^k r_n^{\beta-1}.$$

Thus, from housing market clearing condition 1,

$$\left[L_n^H V^H + L_n^L V^L\right] \beta r_n^{\beta-1} = 1.$$

This last expression appears to suggest that population increases with rents, consistent with Fact 1. However, this is not necessarily true, since wages of low-skilled workers are generally smaller than those of high-skilled workers, that is, $V^L < V^H$. In the Appendix, we show that, for small variations in λ_n around a cross-city average, the population

of both types of workers increases with λ_n if and only if $\beta\theta < 1$. To see why this is necessary, note that as λ_n increases, firms will hire fewer low-skilled workers for each unit of output they produce. How much substitution occurs depends on the elasticity parameter θ. However, higher λ_n also implies higher overall output. Thus, the net effect in the demand for low-skilled workers is ambiguous. Output will increase more with productivity if rents are a small share of wage income (low β). Otherwise, workers will demand large wage increases in order to offset small increases in rent, thus limiting output variation. To summarize, L_n^L increases with λ_n if firms are not too ready to substitute between the two types of workers (i.e., if θ is low) and if wages are not too sensitive to rents (i.e., if β is low). Fortunately, those two parameters have been amply estimated. Typically, the housing share of consumption is pinned at $\beta = 1/3$, and the elasticity of substitution between college- and noncollege-educated workers is smaller than $\theta = 2$.[7] Thus, we can safely assume that $\beta\theta < 1$, so that demand for both types of employment rises with productivity of skilled workers. It follows that variations in λ_n can thus also account for Fact 1.[8]

Finally, in order to explain Fact 3, we need to depart from the model with homogeneous preferences. If workers have identical preferences for living in all cities, the wage premium has to be the same in all cities $(w_n^H/w_n^L = V^H/V^L)$. Fact 3 emerges once one allows for such labor heterogeneity. Manipulating the labor demand and labor supply equations (3), (4), and (6), we thus have that

$$\left(\frac{w_n^H/V^H}{w_n^L/V^L}\right)^{\nu} = \frac{L_n^H}{L_n^L} = \left(\frac{w_n^H}{w_n^L}\right)^{-\theta}(\lambda_n)^{\theta-1},$$

where the first equation we obtain from labor supply and the second from labor demand. Combining the two equations then yields the wage premium:

$$\frac{w_n^H}{w_n^L} = \left(\frac{V^H}{V^L}\right)^{\frac{\nu}{\nu+\theta}}(\lambda_n)^{\frac{\theta-1}{\nu+\theta}}.$$

It follows that, so long as there is some heterogeneity in preferences $(\nu < \infty)$ and $\theta > 1$, the wage premium increases with λ_n.

---

[7] See Ciccone and Peri (2005) for estimates of θ. A value of β close to $1/3$ has been used by Hsieh and Moretti (2017) among others.

[8] One interesting question is what happens if there are multiple tradable industries with different intensities of use for different workers. Then substitution at the city level may be higher than at the firm level or in the aggregate, since it can operate through changes in the industrial composition.

Sources of Variation in Demand for Skilled Labor

The recent literature has identified two main sources of variation in the demand for skilled labor: differential biased technical progress, expressed by differences in adoption of computing technology, and externalities. The first source has been emphasized by Beaudry et al. (2010) and Autor and Dorn (2013), whereas the second has been emphasized by Baum-Snow et al. (forthcoming). We discuss those two in turn.[9] A third source of variation in the demand for skilled labor that has been less explored in the literature comes from the industrial composition of cities and, in particular, the importance of the business services sector (Hendricks 2011). We discuss this third source last.

Computers

One explanation for the increase in the productivity of skilled workers is the adoption of computing technology (Krusell et al. 2000). This is a natural hypothesis, since computers have become cheaper in the same period in which wage inequality has increased. At the same time, in that same period, many of the trends associated with Facts 1, 2, and 3 have been in play. This has motivated Beaudry et al. (2010) and Autor and Dorn (2013) to propose computerization as an explanation for cross-city variation in wage inequality trends.

To see the role of computer adoption, extend the model to allow the production function to include three inputs, the third of which is capital:

$$F\left(L_n^L, L_n^H, K_n\right) = \left[\left(L_n^L\right)^{\frac{\theta-1}{\theta}} + \left(\left(\frac{1}{\alpha}K_n\right)^\alpha \left(\frac{1}{1-\alpha}L_n^H\right)^{1-\alpha}\right)^{\frac{\theta-1}{\theta}}\right]^{\frac{\theta}{\theta-1}},$$
$$\theta > 1$$

Note that capital is complementary to skilled labor but not to unskilled labor. In the spatial setting, Autor and Dorn (2013) have motivated this complementarity through a task-based approach. Computers are particularly adept at enhancing the ability of workers performing abstract tasks involving creativity, coordination, and problem solving, whereas automation can serve as a substitute for workers performing routine tasks such as bookkeeping, clerical work, and repetitive production.

---

[9] Giannone (2017) is a recent contribution analyzing the implications of biased technical progress and local externalities in a common framework.

Suppose computers are a perfectly tradable good, which can be purchased in any city n at the same price p. The price should be understood to incorporate the "quality-adjusted" cost of computers, so that its time variation would include all the large gains in computational power in the past few decades.[10] Then cost minimization implies that $K_n = \frac{\alpha}{1-\alpha}\frac{w_n^H}{p}L_n^H$, so that substituting into the production function we have:

$$Y_n = \left[\left(L_n^L\right)^{\frac{\theta-1}{\theta}} + \left(\frac{1}{1-\alpha}\left(\frac{w_n^H}{p}\right)^\alpha L_n^H\right)^{\frac{\theta-1}{\theta}}\right]^{\frac{\theta}{\theta-1}}.$$

The resulting function on the right-hand side is isomorphic to the one in our basic model, with the labor-augmenting technology parameter substituted for an increasing function of the relative cost of skilled labor and computers $(\lambda_n = \frac{1}{1-\alpha}\left(w_n^H/p\right)^\alpha)$. As the relative price of computers decreases, firms use more of those, increasing the relative productivity of skilled workers.

With homogeneous preferences for location $(\nu \to \infty)$, we can rewrite the production function in terms of rents:

$$Y_n = \left[\left(L_n^L\right)^{\frac{\theta-1}{\theta}} + \left(\frac{1}{1-\alpha}\left(\frac{(r_n)^\beta}{p}V^H\right)^\alpha L_n^H\right)^{\frac{\theta-1}{\theta}}\right]^{\frac{\theta}{\theta-1}}.$$

Cities with higher rents are also cities where wages are higher and where firms have the most incentives to adopt computers. This, in turn, generates a correlation between rents and the labor mix of cities. In the presence of heterogeneity in preferences for different cities $(\nu < \infty)$, cities with higher rents will also feature more wage inequality. Of course, once we endogenize λ_n, we need to have another source of exogenous variation in order to explain differences in rents across cities. One possibility is to still allow for some exogenous variation in the skill premium, possibly due to the location of universities or to allow for citywide productivity differentials.

In order to explain the observed trends, Beaudry et al. (2010) examine an environment in which they can explain observed trends just from the reduction in the price of computers (p). In their environment, apart from the three factors of production, firms have the option to

---

[10] Suppose a unit of an input purchased at a price $\tilde{p}$ produces z units of output according to a linear technology. Then one can define the quality-adjusted value of the input by $p = \tilde{p}/z$. In our example, the output would be computational power, and z would capture the increase in quality of computers.

choose between technologies with higher or lower intensity in skilled labor. As computers become cheaper, the incentive to use a technology that is more intensive in skilled labor increases.

Beaudry et al. (2010) also show, however, that this incentive is stronger in cities where the supply of skilled labor is higher to begin with.[11] This generates a correlation between the adoption of computers and the cross-city supply of skilled labor before computers were adopted, which they verify in the data. Intriguingly, they also show that before computers became an important part of production, the cross-city correlation between skill composition and skill premium was negative, as predicted by our model if there are no important differences in λ_n across cities but there are differences in relative labor supply. As time has progressed and computer adoption has increased, they find that correlation flipping and becoming insignificantly different from zero. One limitation of their explanation for the observed trends is that it cannot account for the positive correlation between skill composition and wage premia observed in recent data (or with other definitions of skill). In their framework, as the price of computers declines, all cities adopt the skill-intensive technology.

Externalities

An alternative explanation for the differences in demand for skilled labor in different cities involves the differential impact of local externalities on different types of workers. Moretti (2004) separates relevant theories into those involving learning and those involving labor markets.[12] Learning-based theories emphasize the role of geographic proximity in facilitating the transfer of knowledge.[13] High-skilled workers perform tasks that are more knowledge-intensive, so they would stand to benefit more from those transfers. Matching theories exploit pecuniary externalities that emerge in imperfect labor markets. A deeper

---

[11] In terms of our model, this would involve extending the utility function to contain a city/skill-specific effect, so that

$$G_n^k(X_n^k, H_n^k, \varepsilon_n(i)) = A_n^k \varepsilon_n(i) \left(\frac{X_n^k}{1-\beta} \right)^{1-\beta} \left(\frac{H_n^k}{\beta} \right)^{\beta}.$$

[12] See Duranton and Puga (2004) for a detailed overview of theories of agglomeration externalities more generally, including the role of sharing, matching, and learning.

[13] In Marshall's (1890) words "Great are the advantages that people following the same trade get from near neighborhood to one another: the mysteries of the trade become no mysteries; but are, as it were, in the air." Lucas (1988) follows up by stating that "Most of what we know we learn from other people (. . .) most of it we get for free. We know that this kind of external effect is common to all arts and sciences – the 'creative professions.'"

pool of skilled workers provides an environment where firms can use them more efficiently.[14]

There is an ample empirical literature pointing to human capital externalities as a source of differences in wages across cities (see Moretti [2004] for a review). This literature focuses on the effects of the skill composition of the workforce. Most recently, the role of externalities in explaining observed trends has been examined by Baum-Snow et al. (forthcoming). In their paper, they focus on what the urban economics literature has called "agglomeration externalities," i.e., external effects associated with overall city size. They allow for city size to affect the productivity of skilled and unskilled labor separately (and also of capital, which they allow for explicitly). They find a robust significant increase over time in the positive agglomeration effect on skilled labor but do not find a robust change in the agglomeration (or congestion) effect on unskilled labor. In terms of our model, they find $\lambda_n = \left(L_n^H + L_n^L \right)^\mu$, where $\mu > 0$.[15] They state that such a change in agglomeration externalities can account for 80 percent of the more rapid increase in wage inequality in large cities, with capital accumulation, the other leading alternative discussed above, accounting for less than 20 percent.

Recent work has also tried to disentangle the two sources of local spillovers. Learning and matching theories have different predictions at the microeconomic level. Learning theories imply that individual productivity is likely to increase with the time that individuals spend in a city and that productivity gains will be embodied in workers who will carry those gains with them if they immigrate. In contrast, matching externalities are unlikely to change with the time a worker has spent in a city and remain specific to the city. The evidence is consistent with learning among the high-skilled workers playing an important role, with Glaeser and Maré (2001), Baum-Snow and Pavan (2013), and De La Roca and Puga (2017) pointing out that high-skilled workers tend to experience faster wage increases when they live in large cities.[16]

Understanding why local externalities have become stronger for skilled workers is challenging. One interesting mechanism is explored by Michaels, Rauch, and Redding (2017). Investigating changes over long periods of time and using a detailed breakdown of occupations into a variety of tasks, they document a rising concentration of occupation-intensive "interactive tasks" in cities. Those tasks are ones "concerned

---

[14] For example, it could reduce the importance of holdout problems, allowing for more investment in skill-complementary capital (Acemoglu 1996).

[15] In reality, they are measuring changes in μ. All the comparative statics presented before also hold for changes in λ_n.

[16] See also Davis and Dingel (2016) for a microfounded spatial model with knowledge spillovers that captures most of the facts above.

with thought, communication, and inter-social activity" and are typically associated with highly skilled occupations such as nurses, accountants, and statisticians. In their model, those tasks tend to become concentrated as the cost of trading them across cities decreases and large cities are able to exploit their comparative advantage in those tasks.

Industrial and Functional Composition

One potential source of variation in demand for skilled labor is variation in industry composition of cities. One extreme case, which is nested in our model, has each city specialize in a single industry. Since industries may have different skill intensities, such industrial specialization would imply cross-city variation in λ_n. For example, Brinkman (2014) points to the increased concentration of certain skill-intensive industries, such as finance, in large cities as an important factor explaining some of the trends discussed. However, as argued by Hendricks (2011), cross-city variation in industrial composition accounts for only a small fraction of cross-city variation in skill composition. Moreover, he shows that most of the variation in skill composition across cities can be tied to variation in a high-skill-specific productivity component that is common to all industries.

At the same time, Hendricks (2011) also finds that cities with a high fraction of skilled workers are also ones with large business services sectors. He proposes a model where, like computers, the output of those services is complementary to high-skilled labor. When external accountants become cheaper, firms hire those to work with their internal staff rather than as a substitute. In terms of our model, cities where business services are cheaper would thus have higher λ_n. He endogenizes the variation in business services productivity by allowing for increasing returns to scale in that sector. One advantage of this focus on the business services sector over a focus on learning or matching externalities is that it provides a mechanism through which the external effects on the employment of skilled workers spread somewhat uniformly over a fairly heterogeneous set of sectors, in consonance with Hendricks's (2011) data analysis.

A special role for the business services sector is shared by Duranton and Puga (2004). They show that this is associated with an increasing specialization of cities by function, meaning that firms have increasingly concentrated executives and managers in larger cities and production activities in smaller ones. They therefore develop a model in which there are gains to concentrating management in cities with a high concentration of business services. While they do not explicitly tie their model to facts about the skill composition of cities and their wage

differentials, it is natural to map workers active in their management function as relatively high-skilled workers and those active in production as relatively low skilled. Given this mapping, their model could also deliver a concentration of high-skilled workers in large cities, as in the data.

Sorting for Unmeasured Skill

Differences in labor demand need not be the only source of the systematic differences in skill composition, wages, and skill premia across cities. An alternative explanation relies on sorting for unobserved worker characteristics. To see how that works, suppose there are multiple types of workers rather than just two, and assume for simplicity that they are all perfect substitutes in production, so that[17]

$$Y_n = \sum_k \mu^k L_n^k.$$

Note that now the production function is the same for all cities. The parameters μ^k capture the marginal product of workers of type k in any city. Since the different types of workers are perfect substitutes, for any worker type k in any city n, it will be the case that $w_n^k = \mu^k$. Observed wage variation across cities can occur if different types of workers are combined in common bins over the course of empirical analysis. For example, a broad "college-educated" group of workers might include workers with some college, with four-year college degrees, and with postgraduate degrees; within each of these bins, workers may have heterogeneous innate abilities. One can then explain many of the empirical facts described in the literature through such sorting (for a detailed exposition see, for example, Davis and Dingel [2017]).

To see how such an effect of skill sorting could come about, reparameterize the utility function of individual workers to:

$$G_n^k(X_n^k, \varepsilon(i)) = \begin{cases} A_n \varepsilon(i) X_n^k \ if \ H_n^k \geq 1 \\ -\infty \ otherwise \end{cases},$$

that is, a worker requires a minimal amount of housing to survive, which we normalize to 1, but beyond that how much housing they purchase is irrelevant to their utility. This is an extreme form of nonhomotheticity

---

[17] There are other ways of generating such sorting. See Eeckhout et al. (2014) for an alternative sorting mechanism based on complementarities between different types of workers in nonhomothetic production technology.

in preferences, with housing considered a "necessity." Note that, in order to show that in this setup it is possible to account for many of the key facts as stemming from differences in labor supply, we now allow for an exogenous city-specific variation in amenities, captured by A_n. Under this parameterization, the value for a worker of type k of living in city n is

$$V_n^k \left(w_n^k, r_n, \varepsilon(i) \right) = \varepsilon(i) A_n \left(w_n^k - r_n \right).$$

Given the Fréchet distribution of $\varepsilon(i)$, the fraction of workers of type k living in city n is

$$\frac{L_n^k}{\bar{L}^k} = \frac{\left(A_n \left(w_n^k - r_n \right) \right)^\nu}{\sum_{n'} \left(A_{n'} \left(w_{n'}^k - r_{n'} \right) \right)^\nu}.$$

As before, the fraction increases with wages and decreases with rents. The difference is that now it increases more rapidly with wages when wages are smaller. Substituting in the labor demand condition $w_n^k = \mu^k$ yields:

$$\frac{L_n^k}{\bar{L}^k} = \frac{\left(A_n \left(\mu^k - r_n \right) \right)^\nu}{\sum_{n'} \left(A_{n'} \left(\mu^k - r_{n'} \right) \right)^\nu}.$$

As rents in a city increase, the labor composition shifts toward workers with higher productivity. Finally, note that since all households buy one unit of housing, housing market equilibrium implies that:

$$\sum_k L_n^k = \bar{H}.$$

Since labor supply is increasing in the city-specific amenity and decreasing in rents, it follows that rents have to increase with the city-specific amenity. Thus, cities with higher amenities also have higher rents and a more skilled workforce. If skill differences are not directly observable, this could translate into higher measured wages for each skill level. If skill differences are particularly hard to observe among the most highly skilled workers, this sorting would translate into a higher wage premium for those cities.

The mechanism described here could be criticized on a priori grounds, since it relies on assuming that housing is a necessity. As a matter of fact, while the share of expenditures on shelter does decrease with income, it decreases by relatively little, with higher-income households spending more dollars. Furthermore, one implication of the sorting mechanism is that more-skilled workers are less sensitive to city-specific

prices. This is contrary to much evidence that finds high-skilled workers to be more mobile.[18]

While such a priori criticism could potentially be accommodated with suitable changes to the model without abandoning some of the key insights, more direct empirical assessments have not been favorable to sorting for unobservable characteristics. A first approach, adopted by Combes, Duranton, and Gobillon (2008), is to assess the effects of sorting on city wage differences through worker fixed effects. Those exercises rely on changes in the wages of workers who migrate between cities to separate the effects of sorting from other city characteristics.[19] While this first approach indicates a large role for sorting, it could underestimate or overestimate those effects to the extent that changes in the wages of migrants are not representative of the differences in wages of the overall population of cities. For example, if individuals migrate from small cities to large cities only when they receive particularly good wage offers, this selection effect would lead to an overestimation of the relationship between city size and wages. On the other hand, if workers move to large cities when they are young, take advantage of local learning opportunities to slowly become better workers, and then move back to smaller cities carrying their new abilities with them, such evidence would tend to underestimate the contribution of large cities to worker productivity since wages would change little at the time of moving. Baum-Snow and Pavan (2012) and De La Roca and Puga (2017) specifically control for these effects and find that allowing for learning within cities is especially important. Once that learning is explicitly allowed for, selection for unobservable skills ceases to be an important source of bias in estimating the effect of city size on productivity.

3.   CONCLUSION

Inequality in the United States has an important spatial component. More-skilled workers tend to reside in larger cities where they earn higher wages. In the meantime, less-skilled workers make lower wages even when they live in those cities. Those relationships appear to have become more pronounced as inequality has increased. The evidence points to externalities among high-skilled workers as a significant contributor to those patterns. This suggests that policy may face an equity efficiency trade-off. The presence of positive externalities among

---

[18] See Notowidigdo (2011) for a discussion of how this fact can be rationalized by allowing for transfers that occur to all households irrespective of their income. This is the mirror image of the mechanism discussed here.

[19] See Combes and Gobillon (2015) for a thorough discussion of these and other econometric issues involved in estimating the impact of city size on wages.

high-skilled workers would imply gains to policies that incentivize them to become more concentrated. However, this would tend to increase inequality across and within cities. As pointed out by Fajgelbaum and Gaubert (2018), who discuss these trade-offs in detail, the inability to separate the place one lives from the place one works may lead a utilitarian planner to choose to redistribute income across cities so as to compensate workers who have high amenity value of living in low-productivity cities. While the aforementioned paper provides a substantial first step in that direction, the literature thoroughly exploring these trade-offs from various angles is still in its infancy.

APPENDIX

Appendix 1: Derivation of Labor Supply and Average Utility with Frechet Distributed Preferences

Under a Fréchet distribution, ε_n has a support between zero and infinity with c.d.f. ε_n is $e^{-\varepsilon_n^{-\nu}}$, and the p.d.f. is $\nu\varepsilon_n^{-\nu-1}e^{-\varepsilon_n^{-\nu}}$. Note that in order for any value x to be smaller than the max of several variables, it must be the case that it is smaller than each one of them. Given that individual draws of ε_n are independent, we can calculate this probability by multiplying the individual c.d.f.'s. Specifically, for any given $\hat\varepsilon_n$.

$$\Pr\left[\hat\varepsilon_n w_n^k/(r_n)^\beta \geq \max_{n'\neq n}\varepsilon_{n'}w_{n'}^k/(r_{n'})^\beta\right] = \prod_{n'\neq n}\Pr\left[\varepsilon_{n'}(w_{n'})^\beta/r_{n'}^k \leq (w_n)^\beta/r_n^k\hat\varepsilon_n\right]$$

$$= \prod_{n'\neq n}\Pr\left[\varepsilon_{n'} \leq \frac{(r_{n'})^\beta/w_{n'}^k}{(r_n)^\beta/w_n^k}\hat\varepsilon_n\right]$$

$$= \prod_{n'\neq n}e^{-\left(\frac{(r_{n'})^\beta/w_{n'}^k}{(r_n)^\beta/w_n^k}\hat\varepsilon_n\right)^{-\nu}}$$

$$= e^{-\Phi_n\hat\varepsilon_n^{-\nu}}$$

where $\Phi_n \equiv \frac{1}{\left(w_n^k/(r_n)^\beta\right)^\nu}\sum_{n'\neq n}\left(w_{n'}^k/(r_{n'})^\beta\right)^\nu$. To find $\Pr\left[\varepsilon_n w_n^k/(r_n)^\beta \geq \max_{n'\neq n}\varepsilon_{n'}w_{n'}^k/(r_{n'})^\beta\right]$ integrate over all $\hat\varepsilon_n$, i.e., calculate

$$\int_0^\infty e^{-\Phi_n\varepsilon^{-\nu}}dF(\varepsilon_n) = \frac{1}{\Phi}\int_0^\infty \Phi\nu\varepsilon_n^{-\nu-1}e^{-\Phi\varepsilon_n^{-\nu}}d\varepsilon_n,$$

where now $\Phi \equiv \frac{\sum_{n'}\left(w_{n'}^k/(r_{n'})^\beta\right)^\nu}{\left(w_n^k/(r_n)^\beta\right)^\nu}$. The integrand is equal to the pdf of a Fréchet with scale parameter Φ and which must therefore integrate to 1. It follows that

$$\int_0^\infty e^{-\Phi_n\varepsilon^{-\nu}}dF(\varepsilon_n) = \Phi^{-1} = \frac{\left(w_n^k/(r_n)^\beta\right)^\nu}{\sum_{n'}\left(w_{n'}^k/(r_{n'})^\beta\right)^\nu}.$$

We can also derive average utility in each city. This is given by $E\left[\varepsilon_n \frac{w_n^k}{r_n^\beta} | \varepsilon_n w_n^k/(r_n)^\beta \geq \max_{n' \neq n} \varepsilon_{n'} w_{n'}^k/(r_{n'})^\beta\right]$. Again, we can write this as

$$\int_0^\infty E\left[\varepsilon_n \frac{w_n^k}{r_n^\beta} | \varepsilon_n \geq \max_{n' \neq n} \varepsilon_{n'} \frac{w_{n'}^k/(r_{n'})^\beta}{w_n^k/(r_n)^\beta}\right] dF(\hat{\varepsilon}_n) = \int_0^\infty \Phi \varepsilon_n \frac{w_n^k}{(r_n)^\beta} \nu \varepsilon_n^{-\nu-1} e^{-\Phi \varepsilon_n^{-\nu}} d\varepsilon_n$$

$$= \frac{w_n^k}{(r_n)^\beta} \int_0^\infty \varepsilon_n \Phi \nu \varepsilon_n^{-\nu-1} e^{-\Phi \varepsilon_n^{-\nu}} d\varepsilon_n$$

$$= \frac{w_n^k}{(r_n)^\beta} \Phi^{\frac{1}{\nu}} \Gamma\left(1 - \frac{1}{\nu}\right)$$

$$= \left[\sum_{n'} \left(w_{n'}^k/(r_{n'})^\beta\right)^\nu\right]^{\frac{1}{\nu}},$$

where the first equality follows from the definition of conditional expectation, and the third equality follows from the fact that the integrand is the expected value of a Fréchet with scale parameter $\Phi^{-\nu}$.

Appendix 2: Comparative Statics with Heterogeneous Preferences

The model can be summarized by the following system of equations:

$$Y_n = \left[\left(L_n^L\right)^{\frac{\theta-1}{\theta}} + \left(\lambda_n L_n^H\right)^{\frac{\theta-1}{\theta}}\right]^{\frac{\theta}{\theta-1}}, \ \theta > 1,$$

$$L_n^L = \left(w_n^L\right)^{-\theta} Y_n,$$
$$L_n^H = (\lambda_n)^{\theta-1} \left(w_n^H\right)^{-\theta} Y_n,$$

$$L_n^k = \left(\frac{w_n^k}{V^k (r_n)^\beta}\right)^\nu \bar{L}^k$$

$$V^k = \left[\sum_{n'} \left(\frac{w_{n'}^k}{(r_{n'})^\beta}\right)^\nu\right]^{\frac{1}{\nu}},$$

$$L_n^L H_n^L + L_n^H H_n^H = 1,$$

$$r_n H_n^k = \beta w_n^k.$$

Consider an equilibrium where all cities have the same λ_n. Log-linearize around that equilibrium to get:

$$dy_n = \eta dl_n^L + (1 - \eta)dl_n^H + (1 - \eta)d\lambda_n, \qquad (8)$$

$$dl_n^L = -\theta dw_n^L + dy_n, \qquad (9)$$

$$dl_n^H = -\theta dw_n^H + dy_n + (\theta - 1)d\lambda_n, \qquad (10)$$

$$dl_n^k = \nu dw_n^k - \nu dv^k - \nu \beta dr_n, \qquad (11)$$

$$dv^k = \sum_{n'} \frac{L_n}{\bar{L}_n}(dw_{n'}^k - \beta dr_{n'}), \qquad (12)$$

$$\sigma(dl_n^L + dh_n^L) + (1 - \sigma)(dl_n^H + dh_n^H) = 0, \qquad (13)$$

$$dr_n + dh_n^k = dw_n^k, \qquad (14)$$

where we use dy_n to denote the log deviation of Y_n from the identical city case and where

$$\sigma = \frac{w_n^L L_n^L}{\sum w_n^k L_n^k},$$

and

$$\eta = \frac{\left(L_n^L\right)^{\frac{\theta-1}{\theta}}}{\left(L_n^L\right)^{\frac{\theta-1}{\theta}} + \left(\lambda_n L_n^H\right)^{\frac{\theta-1}{\theta}}}.$$

We can show that $\sigma = \eta$. To see this, solving the demand equations for w_n^k and multiplying both sides by L_n^k yields

$$w_n^L L_n^L = \left(L_n^L\right)^{\frac{\theta-1}{\theta}} (Y_n)^{\frac{1}{\theta}},$$

$$w_n^H L_n^H = \left(\lambda_n L_n^H\right)^{\frac{\theta-1}{\theta}} (Y_n)^{\frac{1}{\theta}}.$$

Plugging those equations in the formula for σ yields the formulas for η.

Next, use the log-linearized housing demand equation (14) to eliminate dh_n^k from the housing market clearing equation (13) and rearrange to get

$$\sigma(dl_n^L + dw_n^L) + (1 - \sigma)(dl_n^H + dw_n^H) = dr_n.$$

Use the labor supply equation (10) to substitute out dw_n^k from the other equations in the system and rearrange. Define $ddy_n = dy_n - d\bar{y}$,

where $d\bar{y}$ denotes the cross-city average log-change in Y_n and analogously for other variables. The system becomes:

$$ddy_n = \eta ddl_n^L + (1-\eta)ddl_n^H + (1-\eta)dd\lambda_n, \qquad (15)$$

$$ddl_n^L = -\frac{\theta\nu}{\theta+\nu}\beta ddr_n + \frac{\nu}{\theta+\nu}ddy_n, \qquad (16)$$

$$ddl_n^H = -\frac{\theta\nu}{\theta+\nu}\beta ddr_n + \frac{\nu}{\theta+\nu}ddy_n + \frac{\nu(\theta-1)}{\theta+\nu}dd\lambda_n, \qquad (17)$$

$$0 = \sum_{n'}\frac{L_n}{\bar{L}_n}(ddw_{n'}^k - \beta ddr_{n'}), \qquad (18)$$

$$\sigma\frac{1+\nu}{\nu}ddl_n^L + (1-\sigma)\frac{1+\nu}{\nu}ddl_n^H = (1-\beta)\,ddr_n. \qquad (19)$$

Using equations (15), (16), and (17) to substitute out ddl_n^k from equation (19) and using $\sigma = \eta$ yields an expression for rents as a function of productivity.

$$ddr_n/dd\lambda_n = \frac{1+\nu}{1+\nu\beta}(1-\eta) > 0.$$

Manipulating (15), (16), and (17) to obtain an expression ddy_n as a function of ddr_n, we can substitute the expression above to get:

$$ddy_n/dd\lambda_n = (1-\eta)\frac{\nu+1}{1+\nu\beta} > 0.$$

For the limiting case with $\nu \to \infty$, this yields

$$ddy_n = \beta^{-1}(1-\eta)dd\lambda_n.$$

For the other limiting case, with $\nu = 0$

$$ddy_n = (1-\eta)dd\lambda_n.$$

We can also obtain expressions for ddl_n^L and ddl_n^H:

$$ddl_n^L/dd\lambda_n = \frac{\nu}{\theta+\nu}(1-\eta)\frac{1+\nu}{1+\nu\beta}\left[1-\theta\beta\right],$$

$$ddl_n^H/dd\lambda_n = \frac{\nu}{\theta+\nu}\left[(1-\eta)\beta\frac{1+\nu}{1+\nu\beta}\left[\beta^{-1}-1\right] + (\theta-1)\left(1-(1-\eta)\frac{\beta+\nu\beta}{1+\nu\beta}\right)\right] > 0.$$

It follows that $ddl_n^L > 0$ if $1 > \theta\beta$, i.e., if land share is sufficiently small or the two types of labor are not too strongly substitutable. Normal calibrations feature $\beta \simeq \frac{1}{3}$ and $\theta \simeq 2$, so that the condition is satisfied and population grows for both types of workers with λ_n, although most strongly for the high type.

Again, with $\nu \to \infty$ the expressions simplify to

$$ddl_n^L/dd\lambda_n = (1 - \eta)\left[\beta^{-1} - \theta\right],$$
$$ddl_n^H/dd\lambda_n = \left[(1 - \eta)\left[\beta^{-1} - 1\right] + \eta\left(\theta - 1\right)\right] > 0,$$

and we verify that with $\nu = 0$ employment does not change:

$$ddl_n^L/dd\lambda_n = 0,$$
$$ddl_n^H/dd\lambda_n = 0.$$

Calculating real wages, we get:

$$ddw_n^L/dd\lambda_n - \beta ddr_n = \frac{1}{\theta + \nu}(1 - \eta)\frac{1 + \nu}{1 + \nu\beta}\left[1 - \theta\beta\right],$$
$$ddw_n^H/dd\lambda_n - \beta ddr_n = \frac{1}{\theta + \nu}\left[(1 - \eta)\beta\frac{1 + \nu}{1 + \nu\beta}\left[\beta^{-1} - 1\right] + (\theta - 1)\left(1 - (1 - \eta)\frac{\beta + \nu\beta}{1 + \nu\beta}\right)\right].$$

Thus, real wages will increase or decrease in relative terms with relative employment. With $\nu \to \infty$, we have that real wages do not change in relative terms at all. With $\nu = 0$,

$$ddw_n^H - \beta ddr_n = \frac{1}{\theta}(1 - \eta)\left[1 - \theta\beta\right],$$
$$ddw_n^H - \beta ddr_n = \frac{1}{\theta}\left[(1 - \eta)\left[1 - \beta\right] + (\theta - 1)(1 - (1 - \eta)\beta)\right].$$

Finally, we can obtain nominal wages:

$$ddw_n^H/dd\lambda_n = \frac{1+\beta\nu}{\theta+\nu}(1-\eta)\frac{1+\nu}{1+\nu\beta},$$

$$ddw_n^H/dd\lambda_n = \frac{1}{\theta+\nu}\left[(1-\eta)\beta\frac{1+\nu}{1+\nu\beta}\left[\beta^{-1}-1\right]+(\theta-1)\left(1-(1-\eta)\frac{\beta+\nu\beta}{1+\nu\beta}\right)\right]$$
$$+\beta\frac{1+\nu}{1+\nu\beta}(1-\eta),$$

so that both nominal wages increase unambiguously. Taking limits, for $\nu \to \infty$

$$ddw_n^L/dd\lambda_n = 1-\eta.$$
$$ddw_n^H/dd\lambda_n = 1-\eta,$$

whereas for $\nu = 0$

$$ddw_n^L/dd\lambda_n = \frac{1}{\theta}\left(1-\eta\right),$$
$$ddw_n^H/dd\lambda_n = 1-\frac{\eta}{\theta}.$$

Comparative Statics with Housing as a Necessity

With housing as a necessity good as in subsection 2.5, we can derive the labor supply function following the same steps as in Appendix 1 to get:

$$L_n^k = \frac{\left(w_n^k - r_n\right)^\nu}{\sum_{n'}\left(w_{n'}^k - r_{n'}\right)^\nu}\bar{L}^k.$$

It follows that, using the same "deviation" notation as above

$$ddL_n^k = \nu\frac{w^k}{w^k - r}ddw_n^k - \nu\frac{r}{w^k - r}ddr_n.$$

Since all workers buy one unit of housing, the housing market equilibrium equation can be written as:

$$\sum_k L_n^k = \bar{H}_n.$$

Suppose the production function is the same in all cities but housing supply is not. The equations defining the equilibrium of the economy in log-linearized form become:

$$ddy_n = \eta ddl_n^L + (1-\eta)ddl_n^H, \tag{20}$$

$$ddl_n^k = -\theta ddw_n^k + ddy_n \ k \in \{L, H\}, \tag{21}$$

$$ddl_n^k = \nu \frac{w^k}{w^k - r} ddw_n^k - \nu \frac{r}{w^k - r} ddr_n, \ k \in \{L, H\}, \tag{22}$$

$$0 = \sum_{n'} \frac{L_n}{\bar{L}_n} \nu (\frac{w^k}{w^k - r} ddw_n^k - \frac{r}{w^k - r} ddr_n), \tag{23}$$

$$\sigma ddl_n^L + (1-\sigma)ddl_n^H = ddh_n, \tag{24}$$

where now $\sigma = \frac{L_n^L}{L_n^L + L_n^H}$. Use the new labor supply equation (22) to eliminate wages from the labor demand equations and the production function to eliminate output:

$$ddl_n^k = -\frac{\theta}{\nu} \left[\frac{w^k - r}{w^k} ddl_n^k + \frac{r}{w^k} ddr_n \right] + \eta ddl_n^L + (1-\eta)ddl_n^H.$$

Combine those with the production function and substitute out the new housing clearing equilibrium (24) to obtain:

$$\left[\eta + \frac{\theta}{\nu} \frac{w^H - r}{w^H} \right] ddl_n^H - \eta ddl_n^L = -\frac{\theta}{\nu} \frac{r}{w^H} ddr_n,$$

$$-(1-\eta)ddl_n^H + \left[1 - \eta + \frac{\theta}{\nu} \frac{w^L - r}{w^L} \right] ddl_n^L = -\frac{\theta}{\nu} \frac{r}{w^L} ddr_n.$$

Solve the system for ddl_n^L and ddl_n^H as functions of ddr_n. From inspection of the structure of the problem, it follows that:

$$ddl_n^H = -\varphi^H ddr_n$$

$$ddl_n^L = -\varphi^L ddr_n,$$

where $\varphi^H \equiv \frac{\theta}{\nu} \frac{\left[\eta + \frac{\theta}{\nu} \frac{w^L - r}{w^L}\right] \frac{r}{w^H} + \eta \frac{r}{w^L}}{D} > 0$ and $\varphi^L \equiv \frac{\theta}{\nu} \frac{\left[1 - \eta + \frac{\theta}{\nu} \frac{w^H - r}{w^H}\right] \frac{r}{w^L} + (1-\eta)\frac{r}{w^H}}{D} > 0$ with $D \equiv \left[\eta + \frac{\theta}{\nu} \frac{w^H - r}{w^H}\right]\left[1 - \eta + \frac{\theta}{\nu} \frac{w^L - r}{w^L}\right] + (1-\eta)\eta$. Note that $\varphi^H > \varphi^L$ if

$$\frac{\theta}{\nu} \left(\frac{w^H}{w^L} - 1 \right) > (1 - 2\eta) \left(1 + \frac{w^H}{w^L} \right).$$

Substituting back into the housing market clearing equation yields:

$$- \left[\sigma \varphi^H + (1 - \sigma) \varphi^L \right] ddr_n = ddh_n.$$

It follows that interest rates fall with housing supply, and employment in both sectors rises. At the same time, the labor mix will change. So long as $\frac{w^H}{w^L}$ is sufficiently greater than 1, skilled labor will increase more rapidly with housing supply than unskilled labor. At the same time, because the elasticity of labor supply at the city level is not infinity, wages for skilled labor will have to grow faster.

Consider now an environment with three worker types, L, HL, and HH. The production function is now

$$Y_n = \left[\left(L_n^L \right)^{\frac{\theta-1}{\theta}} + \left(\lambda \left(L_n^{HL} + \mu L_n^{HH} \right)^{\frac{\theta-1}{\theta}} \right) \right]^{\frac{\theta}{\theta-1}}.$$

For simplicity, we will focus on the case with $\theta \to \infty$ so that high- and low-skilled workers are also perfect substitutes. Suppose now there are common city-specific amenities, so that their utility of living in city n is now:

$$G_n^k(X_n^k, \varepsilon(i)) = \begin{cases} A_n \varepsilon(i) X_n^k \; if \; H_n^k \geq 1 \\ -\infty \; otherwise \end{cases},$$

where A_n represents amenities such as good public schools or nice weather. Those factors are common for all types of labor. Labor supply for $k \in \{LH, HH\}$

$$L_n^k = \frac{\left(A_n \left(w_n^k - r_n \right) \right)^{\nu}}{\sum_{n'} \left(A_{n'} \left(w_{n'}^k - r_{n'} \right) \right)^{\nu}} \bar{L}^k.$$

Using the fact that different types of labor are perfect substitutes so that their wages do not vary in the cross-section, the log-linearized system in deviation form becomes:

$$ddy_n = (1 - \eta^{LH} - \eta^{HH}) ddl_n^L + \eta^{LH} ddl_n^{LH} + \eta^{HH} ddl_n^{HH}, \qquad (25)$$

$$ddl_n^k = -\nu \frac{r}{w^k - r} ddr_n + \nu dda_n^k, \; k \in \{L, HL, HH\}, \qquad (26)$$

$$0 = \sum_{n'} \frac{L_n}{\bar{L}_n} \nu \frac{r}{w^k - r} ddr_n, \qquad (27)$$

$$(1 - \sigma^{HL} - \sigma^{HH}) ddl_n^L + \sigma^{HL} ddl_n^{HL} + \sigma^{HH} ddl_n^{HH} = 0, \qquad (28)$$

Substitute the new labor supply equations into the market clearing equation to get:

$$\left[(1 - \sigma^{HL} - \sigma^{HH})\frac{r}{w^L - r} + \sigma^{HL}\frac{r}{w^{HL} - r} + \sigma^{HH}\frac{r}{w^{HH} - r}\right] ddr_n = dda_n.$$

It follows that rental rates rise with amenities. At the same time, from equation (26) it follows immediately that employment reacts more to interest rates if wages are smaller. Thus, a given increase in amenities will imply larger increases in the population of workers with higher wages. At the same time, even though wages do not change, if one cannot observe HL and HH type workers separately, one will see their average wage rise, as the composition of high-skilled workers shifts toward the ones with a very high skill. Thus, cities with higher amenities will have a more skilled labor force and higher wage premia. One can further generate the correlation with city size by allowing housing supply to be rent elastic.

REFERENCES

Acemoglu, Daron. 1996. "A Microfoundation for Social Increasing Returns in Human Capital Accumulation." *Quarterly Journal of Economics* 111 (August): 779–804.

Albouy, David. 2012. "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas." Working Paper (May).

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103 (August): 1553–97.

Bacolod, Marigee, Bernardo S. Blum, and William C. Strange. 2009. "Skills in the City." *Journal of Urban Economics* 65 (March): 136–53.

Baum-Snow, Nathaniel, Matthew Freedman, and Ronni Pavan. Forthcoming. "Why Has Urban Inequality Increased?" *American Economic Journal: Applied Economics.*

Baum-Snow, Nathaniel, and Ronni Pavan. 2012. "Understanding the City Size Wage Gap." *Review of Economic Studies* 79 (January): 88–127.

Baum-Snow, Nathaniel, and Ronni Pavan. 2013. "Inequality and City Size." *Review of Economics and Statistics* 95 (December): 1535–48.

Beaudry, Paul, Mark Doms, and Ethan Lewis. 2010. "Should the Personal Computer be Considered a Technological Revolution? Evidence From U.S. Metropolitan Areas." *Journal of Political Economy* 118 (October): 988–1036.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. "Productive Cities: Sorting, Selection, and Agglomeration." *Journal of Political Economy* 122 (June): 507–53.

Berry, Christopher R., and Edward L. Glaeser. 2005. "The Divergence of Human Capital Levels Across Cities." *Papers in Regional Science* 84 (August): 407–44.

Brinkman, Jeffrey. 2014. "The Supply and Demand of Skilled Workers in Cities and the Role of Industry Composition." Federal Reserve Bank of Philadelphia Working Paper 14-32 (October).

Ciccone, Antonio, and Giovanni Peri. 2005. "Long-Run
    Substitutability Between More and Less Educated Workers:
    Evidence from U.S. States, 1950–1990." *Review of Economics and
    Statistics* 87 (November): 652–63.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon.
    2008. "Spatial Wage Disparities: Sorting matters!" *Journal of
    Urban Economics* 63 (March): 723–42.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics
    of Agglomeration Economies." In *Handbook of Urban and
    Regional Economics Vol. 5*, edited by Gilles Duranton, J. Vernon
    Henderson, and William Strange. Amsterdam: North Holland,
    247–348.

Davis, Donald R., and Jonathan I. Dingel. 2016. "A Spatial
    Knowledge Economy." Working Paper (August).

Davis, Donald R., and Jonathan I. Dingel. 2017. "The Comparative
    Advantage of Cities." Working Paper (March).

De La Roca, Jorge, and Diego Puga. 2017. "Learning by working in
    big cities." *Review of Economic Studies* 84 (January): 106–42.

Diamond, Rebecca. 2016. "The Determinants and Welfare
    Implications of US Workers' Diverging Location Choices by Skill:
    1980–2000." *American Economic Review* 106 (March): 479–524.

Duranton, Gilles, and Diego Puga. 2004. "Micro-foundations of Urban
    Agglomeration Economies." In *Handbook of Urban and Regional
    Economics Vol. 4*, edited by J. Vernon Henderson and
    Jacques-François Thisse. Amsterdam: Elsevier, 2063–117.

Duranton, Gilles, and Diego Puga. 2005. "From Sectoral to
    Functional Urban Specialisation." *Journal of Urban Economics* 57
    (March): 343–70.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny. 2014.
    "Spatial Sorting." *Journal of Political Economy* 122 (June):
    554–620.

Fajgelbaum, Pablo D., and Cecile Gaubert. 2018. "Optimal Spatial
    Policies, Geography and Sorting." Working Paper 24632.
    Cambridge, Mass.: National Bureau of Economic Research. (May).

Giannone, Elisa. 2017. "Skilled-Biased Technical Change and Regional
    Convergence." University of Chicago Working Paper (January).

Glaeser, Edward L., and David C. Maré. 2001. "Cities and Skills."
    Journal of Labor Economics 19 (April): 316–42.

Handbury, Jessie, and David E. Weinstein. 2015. "Goods Prices and Availability in Cities." *Review of Economic Studies* 82 (January), 258–96.

Hendricks, Lutz. 2011. "The Skill Composition of U.S. Cities." *International Economic Review* 52 (February): 1–32.

Hsieh, Chang-Tai, and Enrico Moretti. 2017. "Housing Constraints and Spatial Misallocation." Working Paper (May).

Krusell, Per, et al. 2000. "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis." *Econometrica* 68 (September): 1029–53.

Lucas Jr., Robert E. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22 (July): 3–42.

Marshall, Alfred. 1890. *Principles of Economics*, 1st ed. London: Macmillan.

Michaels, Guy, Ferdinand Rauch, and Stephen Redding. 2017. "Task Specialization in U.S. Cities from 1880-2000." Working Paper 18715. Cambridge, Mass.: National Bureau of Economic Research. (August).

Moretti, Enrico. 2004. "Human Capital Externalities in Cities." In *Handbook of Urban and Regional Economics Vol. 4*, edited by J. Vernon Henderson and Jacques-François Thisse. Amsterdam: Elsevier, 2243–91.

Moretti, Enrico. 2013. "Real Wage Inequality." *American Economic Journal: Applied Economics* 5 (January): 65–103.

Notowidigdo, Matthew J. 2011. "The Incidence of Local Labor Demand Shocks." Working Paper 17167. Cambridge, Mass.: National Bureau of Economic Research. (December).

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (December): 1257–78.

Rosen, Sherwin. 1979. "Wage-Based Indexes of Urban Quality of Life." In *Current Issues in Urban Economics Vol. 3*, edited by Peter Mieszkowski and Mahlon Straszheim. Baltimore: Johns Hopkins University Press, 324–45.