CHAPTER *14*

Local Labor Markets☆

Enrico Moretti
UC Berkeley, NBER, CEPR and IZA

Contents

Abstract

I examine the causes and the consequences of differences in labor market outcomes across local labor markets within a country. The focus is on a long-run general equilibrium setting, where workers and firms are free to move across localities and local prices adjust to maintain the spatial equilibrium. In particular, I develop a tractable general equilibrium framework of local labor markets with heterogenous labor. This framework is useful in thinking about differences in labor market outcomes of different skill groups across locations. It clarifies how, in spatial equilibrium, localized shocks to a part of the labor market propagate to the rest of the economy through changes in employment, wages and local prices and how this diffusion affects workers' welfare. Using this framework, I address three related questions. First, I analyze the welfare consequences of productivity differences across local labor markets. I seek to understand what happens to the wage, employment and utility of workers with different skill levels when a local economy experiences a shift in the productivity of a group of workers. Second, I analyze the causes of productivity differences across local labor markets. To a large extent, productivity differences within a country are unlikely to be exogenous. I review the theoretical and empirical literature on agglomeration economies, with a particular focus on studies that are relevant for labor economists. Finally, I discuss the implications for policy.

Keywords: Cities; Wage; General equilibrium; Spatial equilibrium

1. INTRODUCTION

Local labor markets in the US are characterized by enormous differences in worker earnings, factor productivity and firm innovation. The hourly wage of workers located in metropolitan areas at the top of the wage distribution is more than double the wage of observationally similar workers located in metropolitan areas at the bottom of the distribution. These differences reflect, at least in part, variation in local productivity. For example, total factor productivity of manufacturing establishments in areas at the top of the TFP distribution is three times larger than total factor productivity in areas at the bottom of the distribution. The amount of innovation is also spatially uneven. Firms in Santa Clara and San Jose generate respectively 3390 and 1906 new patents in a typical year, while the median city generates less than 1 patent per year. Notably, these differences in wages, productivity and innovation appear to be largely persistent over the last three decades.

In this chapter, I review what we know about the causes and the consequences of differences in labor market outcomes across local labor markets within a country. The focus is on a long-run general equilibrium setting, where workers and firms are free to move across localities and local prices adjust to maintain the spatial equilibrium. In particular, I develop a tractable general equilibrium framework of local labor markets

with heterogenous labor. This framework—which represents the unifying theme of the chapter—is useful in thinking about differences in labor market outcomes of different skill groups across locations. It clarifies how, in spatial equilibrium, localized shocks to a part of the labor market propagate to the rest of the economy through changes in employment, wages and local prices and how this diffusion affects workers' welfare.

Using this framework, I address three related questions.

1. First, I analyze the welfare consequences of productivity differences across local labor markets. I seek to understand what happens to the wage, employment and utility of workers with different skill levels when a local economy experiences a shift in the productivity of a group of workers. I focus on welfare incidence and use the spatial equilibrium model to clarify who ultimately benefits from permanent productivity shocks.
2. Second, I analyze the causes of productivity differences across local labor markets. To a large extent, productivity differences within a country are unlikely to be exogenous. I review the theoretical and empirical literature on agglomeration economies, with a particular focus on studies that are relevant for labor economists.
3. Finally, I discuss the implications for policy, with a special focus on location–based economic development policies aimed at creating local jobs. I clarify when these policies are wasteful, when they are efficient and who the expected winners and losers are.

The topic of local labor markets should be of great interest to labor economists for two reasons. First, the issue of localization of economic activity and its effects on workers' welfare is one of the most exciting and promising research grounds in the field. This area, at the intersection of labor and urban economics, is ripe with questions that are both of fundamental importance for our understanding of how labor markets operate and have deep policy implications. Why are some cities prosperous while others are not? Given that factors of production can move freely within a country, why do firms locate in expensive labor markets? What are the ultimate effects of these differences on workers' welfare? These questions have intrigued economists for more than two centuries, but it is only in the last three decades that a body of high quality empirical evidence has begun to surface. The pace of empirical research in this area has accelerated in the last 10-15 years. It is a topic whose relative importance within the field of labor economics promises to keep growing in the next decade.

Second, and more generally, the issue of equilibrium in local labor markets should be of broader interest for all labor economists, even those who are not directly interested in economic geography per se. With notable exceptions, labor economists have traditionally approached the analysis of labor market shocks using a partial equilibrium analysis. However, a partial equilibrium analysis misses important parts of the picture, since the endogenous reaction of factor prices and quantities can significantly alter the ultimate effects of a shock. Because aggregate shocks to the labor market are

rarely geographically uniform, the geographic reallocation of factors and local price adjustments are empirically important. It is difficult to fully understand *aggregate* labor market changes—like changes in relative wages or employment—if ignoring the spatial dimension of labor markets. Partial equilibrium analyses can be particularly misleading in the case where the workforce is highly mobile, like in the US. Labor flows across localities and changes in local prices have the potential to undo some of the direct effects of labor market shocks. This can profoundly alter the implications for policy. In this respect, the workings of local labor markets and their spatial equilibrium cannot be overlooked by labor economists, even those who are working on more traditional topics like wage determination, wage inequality or unemployment.

As an example, consider a nationwide increase in the productivity of skilled workers in an industry, say the software industry. Although the shock is nationwide, it affects different local labor markets differently because the software industry—like most industries—is spatially concentrated. The effect on the demand for skilled labor in a city like San Jose–in the heart of Silicon Valley—is likely to be quite different from the effect in a city like Phoenix—where the software sector is nonexistent. In a partial equilibrium setting, the only effect of this shock is an increase in the nominal wage of skilled workers in San Jose. However, in general equilibrium this shock propagates to other parts of the economy through changes in factor prices and quantities. Indeed, in general equilibrium, *all* agents in the economy are affected, irrespective of their location and their skill level. Attracted by higher demand, some skilled workers leave Phoenix and move to San Jose, thus pushing up the cost of housing and other non-tradable goods there. Unskilled workers in San Jose are affected because cost of living increases and because of imperfect substitution between skilled and unskilled labor. On net, some unskilled workers move to Phoenix, attracted by higher real wages. Skilled and unskilled workers in Phoenix also experience changes in their equilibrium wage, even if their productivity has not changed, because of changes in their local supply. Following population changes, owners of land experience changes in the value of their asset, both in San Jose and Phoenix. In this example, the direct effect of the demand shock is partially offset by general equilibrium changes due to worker relocation and local price adjustments. The ultimate change in the nominal and real wage of skilled and unskilled workers—and their policy implications—are quite different from the partial equilibrium change and crucially depends on the degree of labor mobility and the magnitude of local prices changes.

The chapter proceeds as follows. I begin by reviewing some important facts on differences in economic outcomes across local labor markets (Section 2). I focus on differences in nominal wages, real wages, productivity and innovation across US metropolitan areas.

I then present the spatial equilibrium model of the labor market (Section 3). The model is kept deliberately very simple, so that all the equilibrium outcomes have easy-to-interpret closed-form solutions. At the same time, the model is general enough to capture many key features of a realistic spatial equilibrium. While there are several versions of the spatial equilibrium model in the literature, and its basic insights are generally well understood, the focus on welfare incidence is relatively new.

In general equilibrium, a shock to a local labor market is partially capitalized into housing prices and partially reflected in worker wages. While marginal workers are always indifferent across locations, the utility of inframarginal workers can be affected by localized shocks. The model clarifies that the welfare consequences of localized productivity shifts depend on which of the two factors of production—labor or housing—is supplied more elastically at the local level.[1] A lower local elasticity of labor supply implies that a larger fraction of a shock to a city accrues to workers in that city and a smaller fraction accrues to landowners in that city. On the other hand, a more inelastic housing supply implies a larger incidence of the shock on landowners, holding constant labor supply elasticity. This makes intuitive sense: if labor is relatively less mobile, local workers are able to capture more of the economic rent generated by the shock. Additionally, a lower local elasticity of labor supply implies a smaller effect on the utility of workers in non–affected cities, since what links different local labor markers is the potential for worker mobility. The model also clarifies how the elasticity of local labor supply is ultimately governed by workers' preferences for location.

A particularly interesting case is what happens when there are two skill groups and one group experiences a localized productivity shock. This question is relevant because skill-specific shocks are common and have important consequences for nationwide inequality. The model clarifies how the relative elasticity of labor supply of different skill groups governs the ultimate effect of the shock on the utility of workers in each skill group and in each city.

Having clarified the welfare consequences of productivity differences across local labor markets, I turn to the possible causes of these differences. Because labor and land costs vary so much across local labor markets, economists have long suspected that there must exist significant productivity differences to offset the differences in factor costs, especially for industries that produce tradable goods. In the absence of significant productivity advantages, why would firms that produce tradable goods be willing to locate in places like Silicon Valley, New York or Boston, which are characterized by exorbitant labor and land costs, rather than in rural areas or in poorer cities, which are characterized by lower factor prices? Ever since at least Marshall (1890), economists have posited that these productivity advantages are not exogenous and may be explained by the existence of agglomeration economies. In Section 4, I review the existing empirical evidence on agglomeration economies, focusing on papers that are particularly relevant to labor economists. I address two related questions. First, what do we know about the magnitude of agglomeration economies? Second, what do we know about the micro mechanisms that generate agglomeration economies? The past two decades have seen a significant amount of effort devoted to answering these questions. Overall, there seems to be growing evidence that points to the fact that in many tradable goods productions, a firm's productivity is higher when it locates close to other similar firms. Notably, these productivity advantages seems to be increasing not only in geographic proximity but also in economic proximity. For example, they are larger for pairs of firms that share similar

---

[1] Capital is assumed to be supplied with infinite elasticity at a price determined by the international market.

labor pools, similar technologies, and similar intermediate inputs. The exact mechanism that generates these economies of scale remains more elusive. I discuss the most important explanations that have been proposed and the empirical evidence on each of them. I conclude that much remains to be done in terms of empirically understanding their relative importance.

Finally, in Section 5, I discuss the efficiency and equity rationales for local development policies aimed at creating local jobs. In the US, state and local governments spend $30-40 billion per year on these policies, while the federal government spends $8-12 billion. While these policies are pervasive, their economic rationale is often misunderstood by the public and economists alike. From the equity point of view, location-based policies aim at redistributing income from areas with high level of economic activity to areas with low level of economic activity. In this respect, these policies are unlikely to be effective. The spatial equilibrium model clarifies that in a world where workers are mobile, local prices adjust so that workers are unlikely to fully capture the benefits of location-based subsidies. When mobility is more limited, these policies have the potential to affect the utility of inframarginal workers', but in ways that are non transparent and difficult to know in advance, because they depend on individual idiosyncratic preferences for location. From an efficiency point of view, the main rationale for these type of subsidies is the existence of significant agglomeration externalities. If the attraction of new businesses to a locality generates localized productivity spillovers, then the provision of subsidies may be able to internalize the externality. The magnitude of the optimal subsidy depends on the exact shape of Marshallian dynamics. In this case, government intervention may be efficient from the point of view of a locality, although not necessarily from the point of view of aggregate welfare.

Ever since Adam Smith wrote his treatise on the "Nature and Causes of the Wealth of Nations" more than two centuries ago, economists have sought to understand the underlying causes of income disparities across regions of the world. While historically economists have focused on understanding the causes of differences *across countries*, the question of differences across localities *within a country* is receiving growing attention. Within county differences in productivity and wages are possibly even more remarkable than cross-country differences, since the mobility of labor and capital within a country is unconstrained and differences in institutions and regulations are small relative to cross-country differences. As a consequence, it is difficult to understand why some countries are poor and other countries are rich without first understanding why some cities within a country are poor and others are rich. The issue of local labor markets is a central one for economists, and much remains to be done to fully understand it.

2. SOME IMPORTANT FACTS ABOUT LOCAL LABOR MARKETS

Most countries in the world are characterized by significant spatial heterogeneity in economic outcomes. As an example, Fig. 1 shows the amount of income produced by square mile in the United States. The map documents enormous differences in the
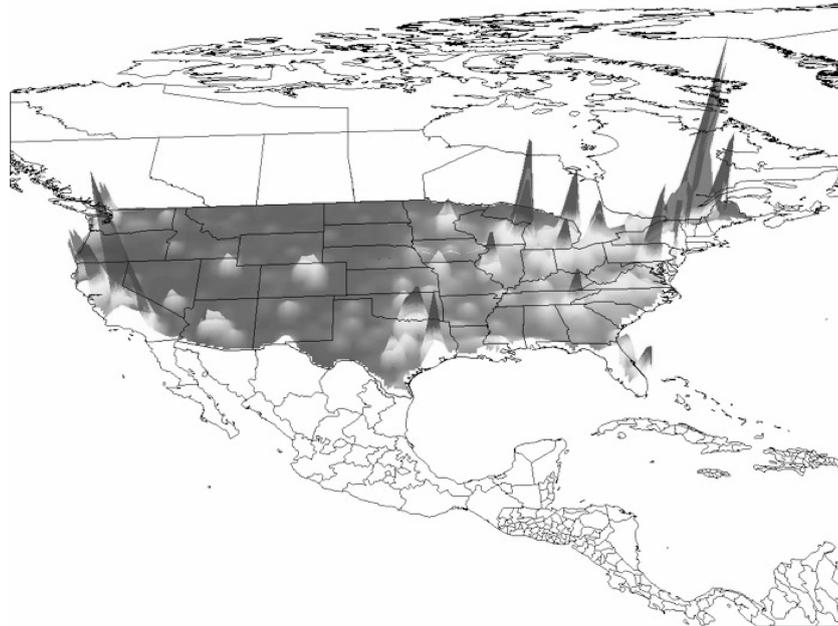
Figure 1   *Spatial distribution of economic output in the US, by square mile.* Notes: This figure reports the value of output produced in the US by square mile.

density of economic activity across different parts of the country. In the US there are a limited number of cities producing most of the country's output, surrounded by vast areas generating little output. Many other developed and developing countries show a similar pattern in the distribution of economic activity.

In this Section, I document the magnitude of the differences in labor market outcomes across local labor markets in the United States. In particular, I focus on spatial differences in nominal wages, real wages, productivity and innovation and how these differences have evolved over the last three decades.[2]

2.1. Nominal wages

The vast differences in output per mile in Fig. 1 translate into equally vast differences in workers' wages. The top panel in Fig. 2 shows the distribution of average hourly nominal wage for high school graduates by metropolitan statistical areas (MSA). Data are from the 2000 Census of Population and include all full-time US workers between the age of 25 and 60 who worked at least 48 weeks in the previous year. The figure

---

[2] Another notable feature of the spatial distribution of economic activity is represented by industry clustering, whereby firms tend to cluster near other "similar" firms (for example: firms that sell similar products). The cluster of IT firms in Silicon Valley, biomedical research in Boston, biotech in San Diego and San Francisco, financial firms in Wall Street and London are notable examples. In this section, I do not focus on this feature. However, I discuss its causes and consequences in the following sections.
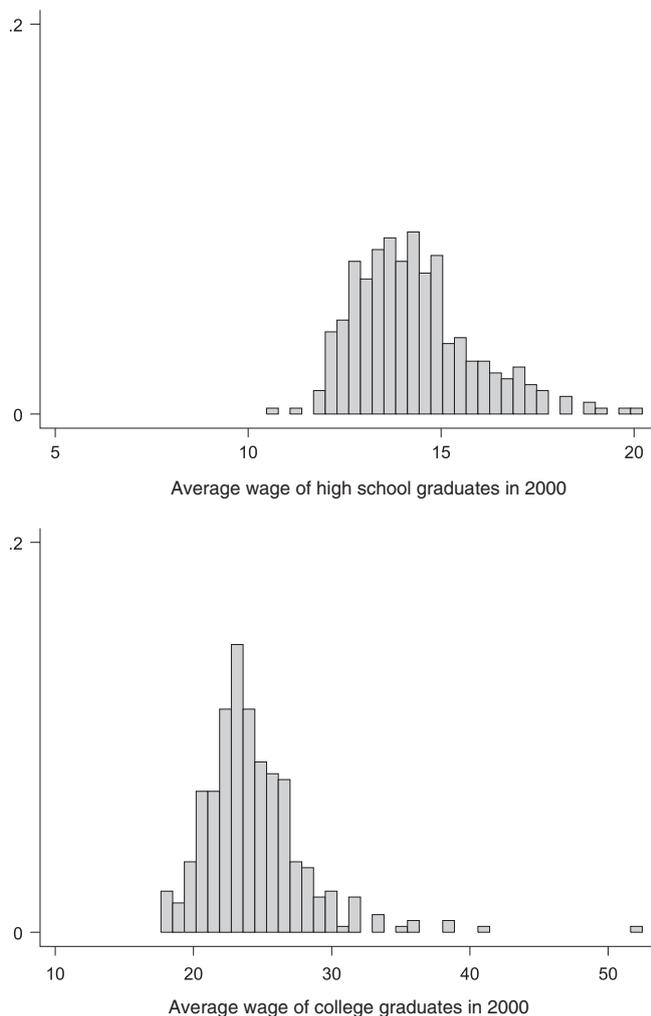
Figure 2 *Distribution of average hourly nominal wage of high school graduates and college graduates, by metropolitan area.* Notes: This figure reports the distribution of average hourly nominal wage of high school graduates and for college graduates across metropolitan areas in the 2000 Census of Population. There are 288 metropolitan areas. The sample includes all full-time US born workers between the age of 25 and 60 who worked at least 48 weeks in the previous year.

indicates that labor costs vary significantly across US metropolitan areas. The average high school graduate living in the median metropolitan area earns $14.1 for each hour worked. The 10th and 90th percentile of the distribution across metropolitan areas are $12.5 and $16.5, respectively. This amounts to a 32% difference in labor costs. The 1st and 99th percentile are $11.9 and $19.0, respectively, which amounts to a 60% difference. While some of these differences may reflect heterogeneity in skill levels within education group, differences across metropolitan areas conditional on race, experience, gender, and Hispanic origin are equally large.

Table 1 Metropolitan areas with the highest and lowest hourly wage of high school graduates in 2000.

Metropolitan area	Average hourly wage
Metropolitan areas with the highest wage	
Stamford, CT	20.21
San Jose, CA	19.70
Danbury, CT	19.13
San Francisco-Oakland-Vallejo, CA	18.97
New York-Northeastern NJ	18.86
Monmouth-Ocean, NJ	18.30
Santa Cruz, CA	18.24
Santa Rosa-Petaluma, CA	18.23
Ventura-Oxnard-Simi Valley, CA	17.72
Seattle-Everett, WA	17.71
Metropolitan areas with the lowest wage	
Ocala, FL	12.12
Dothan, AL	12.11
Amarillo, TX	12.10
Danville, VA	12.08
Jacksonville, NC	12.02
Kileen-Temple, TX	11.98
El Paso, TX	11.96
Abilene, TX	11.87
Brownsville-Harlingen-San Benito, TX	11.23
McAllen-Edinburg-Pharr-Mission, TX	10.65

The sample includes all full-time US born workers between the age of 25 and 60 with a high school degree who worked at least 48 weeks in the previous year. Data are from the 2000 Census of Population.

The bottom panel in Fig. 2 shows the distribution of average hourly nominal wage for college graduates across metropolitan areas. (Note that the scale in the two panels is different.) The distribution of the average wage of college graduates across metropolitan areas is even wider than the distribution of the average wage of high school graduates. The 10th and 90th percentile of the distribution for college graduates are $20.5 and $28.5. This amounts to a 41% difference in labor costs. The 1st and 99th percentile are $18.1 and $38.5, respectively, which amounts to a 112% difference.

Table 1 lists the 10 metropolitan areas with the highest average wage for high school graduates and the 10 metropolitan areas with the lowest average wage for high school graduates. High school graduates living in Stamford, CT or San Jose, CA earn an hourly wage that is two times as large as workers living in Brownsville, TX or McAllen, TX with the same level of schooling. This difference remains effectively unchanged after accounting for differences in workers' observable characteristics. Table 2 produces a similar list for college graduates. The difference between wages in cities at the top of the distributions and cities at the bottom of the distribution is more pronounced for

Table 2   Metropolitan areas with the highest and lowest hourly wage of college graduates in 2000.

Metropolitan area	Average hourly wage
Metropolitan areas with the highest wage	
Stamford, CT	52.46
Danbury, CT	40.81
Bridgeport, CT	38.82
San Jose, CA	38.49
New York–Northeastern NJ	36.03
Trenton, NJ	35.52
San Francisco-Oakland-Vallejo, CA	34.89
Monmouth-Ocean, NJ	33.70
Los Angeles-Long Beach, CA	33.37
Ventura-Oxnard-Simi Valley, CA	33.07
Metropolitan areas with the lowest wage	
Pueblo, CO	20.16
Goldsboro, NC	20.15
St. Joseph, MO	20.01
Wichita Falls, TX	19.74
Abilene, TX	19.70
Sumter, SC	19.57
Sharon, PA	19.52
Waterloo-Cedar Falls, IA	18.99
Altoona, PA	18.68
Jacksonville, NC	18.21

The sample includes all full-time US born workers between the age of 25 and 60 with a college degree who worked at least 48 weeks in the previous year. Data are from the 2000 Census of Population.

college graduates. The average hourly wage of college graduates in Stamford, CT is almost three times larger than the hourly wage of college graduates in Jacksonville, NC. This difference is robust to controlling for worker characteristics.

The wage differences documented in Fig. 2 are persistent over long periods of time. While in the decades after World War II regional differences in income were declining (Barro and Sala-i-Martin, 1991), convergence has slowed down significantly in more recent decades. This can be seen in Fig. 3, where I plot the average hourly wage in 1980 against the average wage in 2000 for high school graduates and college graduates, by metropolitan area. The size of the bubbles is proportional to the number of workers in the relevant metropolitan area and skill group 1980. The lines are the predicted wages in 2000 from a weighted OLS regression, where the weights are the number of workers in the relevant metropolitan area and skill group in 1980.

The figure suggests that there has been no mean reversion in wages since 1980. In fact, the opposite has happened. Wage differences across metropolitan areas have *increased* over time. The slope of the regression line is 1.82 (0.89) for high school graduates. This

Figure 3   *Change over time in the average hourly nominal wage of high school graduates and college graduates, by metropolitan area.* Notes: Each panel plots the average nominal wage in 1980 against the average nominal wage in 2000, by metropolitan area. The top panel is for high school graduates. The bottom panel is for college graduates. The size of the bubbles is proportional to the number of workers in the relevant metropolitan area and skill group 1980. There are 288 metropolitan areas. The line is the predicted wage in 2000 from a weighted OLS regression, where the weights are the number of workers in the relevant metropolitan area and skill group in 1980. The slope is 1.82 (0.89) for high school graduates and 3.54 (0.11) for college graduates. Data are from the Census of Population. The sample includes all full-time US born workers between the age of 25 and 60 who worked at least 48 weeks in the previous year.

Table 3 Average hourly wage in 1980 and 2000, by education level and metropolitan area.

High school graduates

	Low wage in 2000	High wage in 2000
Low wage in 1980	106	40
High wage in 1980	34	108

College graduates

	Low wage in 2000	High wage in 2000
Low wage in 1980	114	32
High wage in 1980	26	116

For each skill group, metropolitan areas are classified as having a low or high wage depending on whether their average wage is below or above the average wage of the median metropolitan area in the relevant year. The sample includes all full-time US born workers between the age of 25 and 60 who worked at least 48 weeks in the previous year. Data are from the 2000 Census of Population. There are 288 metropolitan areas.

suggests that metropolitan area where high school graduates have high wages in 1980 compared to other metropolitan areas have even higher wages in 2000. The slope for college graduates is 3.54 (0.11). The fact that the slope is even higher for college graduates indicates that the increase in the spatial differences in hourly wages is larger for skilled workers.

The lack of spatial convergence is also documented in Table 3, where I classify metropolitan areas as having low or high wage depending on whether the average wage is below or above the average wage in the median metropolitan area in the relevant year. This is done separately for each year and each education group. The top panel shows that in most cases, metropolitan areas where high school graduates have high wages in 1980 also have high wages in 2000. Only a quarter of metropolitan areas change category. Consistent with the larger increase in spatial divergence uncovered in Fig. 3, this fraction is even smaller for college graduates (bottom panel).

Using data on total income instead of hourly wages, Glaeser and Gottlieb (2009) find no evidence of convergenece across metropolitan areas between 1980 and 1990, but they find some evidence of convergence between 1990 and 2000. The difference between their findings and Fig. 3 is explained by three factors. First, I am interested in labor market outcomes, so that my sample includes only workers. By contrast, the Glaeser and Gottlieb sample includes all individuals. Second, there may be differences across metropolitan areas in unearned income. Third, and most importantly, there might be differences across metropolitan areas in number of hours worked, since it is well known that, since 1980, workers with high nominal wages have experienced relatively larger increases in number of hours worked than workers with low nominal wages. The convergence in total income uncovered by Glaeser and Gottlieb (2009) in the 1990s is quantitatively limited. Consistent with my interpretation of Fig. 3, they conclude that

although there has been some convergence in income, over the last three decades "rich places have stayed rich and poor places have stayed poor".

When thinking about localization of economic activity, nominal wages are more important than income because they are related to labor productivity. Since labor, capital and goods can move freely within a country, it is difficult for an economy in a long-run equilibrium to maintain significant spatial differences in nominal labor costs in the absence of equally large productivity differences. Indeed, if labor markets are perfectly competitive, nominal wage differences across local labor markets should exactly reflect differences in the marginal product of labor in industries that produce tradable goods. If this were not the case, firms in the tradable sector located in cities with nominal wages higher than labor productivity would relocate to less expensive localities. While not all workers are employed in the tradable sector, as long as there are some firms producing traded goods in every city and workers can move between the tradable and non-tradable sector, average productivity has to be higher in cities where nominal wages are higher.

Overall, if wages are related to marginal product of labor, there appears to be little evidence of convergence in labor productivity across US metropolitan areas. If anything, there is evidence of divergence: metropolitan areas that are characterized by high labor productivity in 1980 are characterized by even higher productivity in 2000. Notably, both the magnitude of geographic differences and speed of divergence appear to be more pronounced for high-skilled workers than low-skilled workers.

2.2. Real wages

The large differences in nominal wages documented above do not appear to be associated with massive migratory flows of workers across metropolitan areas.[3] The main reason for the lack of significant spatial reallocation of labor is that land prices vary significantly across locations so that differences in real wages are significantly smaller than differences in nominal wages. Figure 4 shows the distribution of average hourly real wage for high school and college graduates across metropolitan areas. Real wages are calculated as the ratio of nominal wages and a local CPI that reflects differences in the cost of housing across locations. The index is described in detail in Moretti (forthcoming). A comparison with Fig. 2 indicates that the distribution of real wages is significantly more compressed than the distribution of nominal wages. For example, the 10th and 90th percentile of the distribution for high school graduates are $10.0 and $11.7. This is only a 17% difference. The 10th and 90th percentile of the distribution for college graduates are $16.7 and $20.4, a 22% difference.

If nominal wages adjust fully to reflect cost of living differences, and if amenity differences are not too important, a regression of log nominal wage on log cost of housing should yield a coefficient approximately equal to the share of income spent on housing

---

[3] In a recent review of the evidence, Glaeser and Gottlieb (2008) conclude that "there has been little tendency for people to move to high income areas".
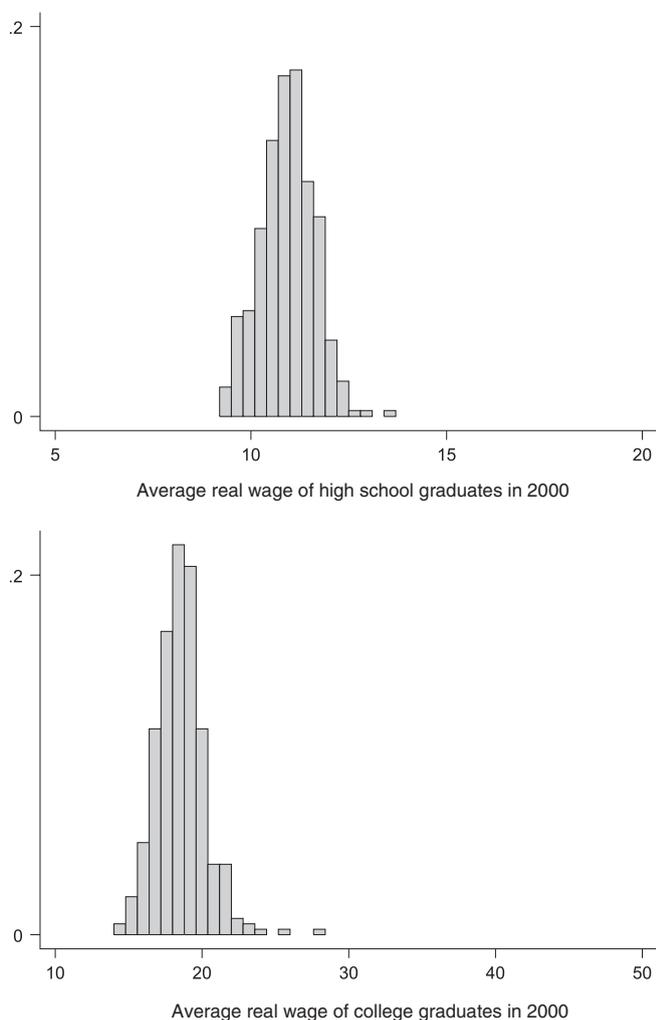
Figure 4   *Distribution of average hourly real wage of high school graduates and college graduates, by metropolitan area.* Notes: This figure reports the distribution of average hourly real wage of high school graduates and college graduates across metropolitan areas in the 2000 Census of Population. Real wage is defined as the ratio of nominal wage and a cost of living index that reflects differences across metropolitan areas in the cost of housing. The index is normalized so that it has a mean of 1. There are 288 metropolitan areas. The sample includes all full-time US born workers between the age of 25 and 60 who worked at least 48 weeks in the previous year.

(Glaeser and Gottlieb, 2008). Empirically, I find that an individual level regression of log earnings on average cost of housing in the metropolitan area of residence—measured by the log average cost of renting a two or three bedroom apartment—controlling for standard observables and clustering the standard errors by metropolitan area yields a
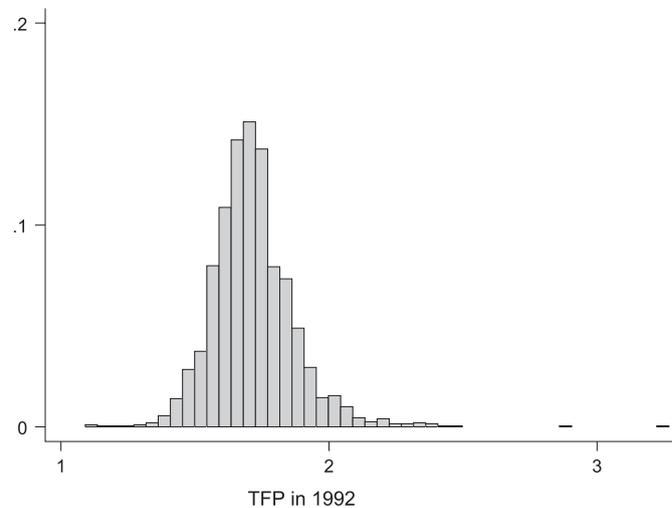
Figure 5 *Distribution of total factor productivity in manufacturing establishments, by county.*
Notes: This figure reports the distribution of average total factor productivity of manufacturing establishments in 1992, by county. County-level TFP estimates are obtained from estimates of establishment level production functions based on data from the Census of Manufacturers. Specifically, they are obtained from a regression of log output on hours worked by blue and white collar workers, book value of building capital, book value of machinery capital, materials, industry and county fixed effects. The figure shows the distribution of the coefficients on the county dummies. Regressions are weighted by plant output. The sample is restricted to counties that had 10 or more plants in either 1977 or 1992 in the 2xxx or 3xxx SIC codes. There are 2126 counties that satisfy the sample restriction. For confidentiality reasons, any data from counties whose output was too concentrated in a small number of plants are not in the figure (although they are included in the regression).

coefficient equal to 0.513 (0.024).[4] Given that the share of income spent on housing is about 41% in 2000, this regression lends credibility to the notion that nominal wages adjust to take into account differences in the cost of living across localities.

2.3. Productivity
The vast differences in nominal wages across local labor markets reflect, at least in part, differences in productivity. Productivity is notoriously difficult to measure directly. One empirical measure of productivity at the establishment level is total factor productivity (TFP), defined as output after controlling for inputs.

Figure 5 shows the distribution of average total factor productivity of manufacturing establishments in 1992 by county. County-level TFP estimates are obtained from estimates of production functions based on data from the Census of Manufacturers. Specifically, they are obtained from a regression of log output on hours worked by blue

---

[4] Data are from the 2000 Census of Population.
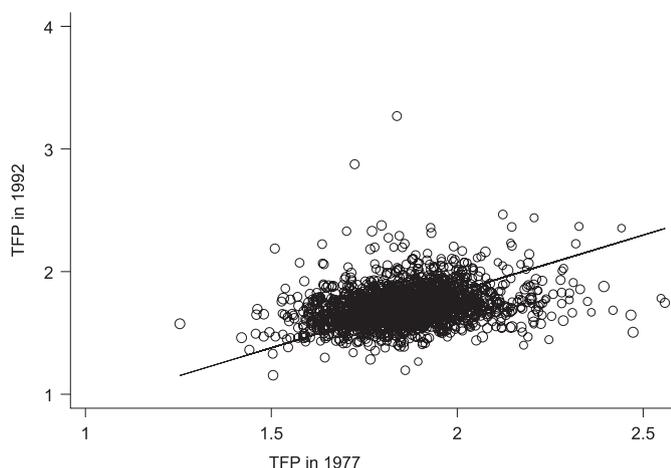
Figure 6   *Change over time in total factor productivity in manufacturing establishments, by county.*
Notes: The figure plots county-level average TFP in 1977 on the x-axis against TFP in 1992 on the y-axis. County-level TFP estimates are obtained from estimates of establishment level production functions based on data from the Census of Manufacturers. Specifically, they are obtained from a regression of log output on hours worked, book value of building capital, book value of machinery capital, materials, industry and county fixed effects. Each regression is estimated separately for 1977 and 1992. The figure shows the coefficients on the county dummies in each year. Regressions are weighted by plant output. The sample is restricted to counties that had 10 or more plants in either 1977 or 1992 in the 2xxx or 3xxx SIC codes. There are 2126 counties that satisfy the sample restriction. For confidentiality reasons, any data from counties whose output was too concentrated in a small number of plants are not in the figure (although they are included in the regression).

collar and white collar workers, building capital, machinery capital, materials, industry fixed effects and county fixed effects.[5] The level of observation is the establishment. The coefficients on the county dummies represent county average total factor productivity, holding constant industry, capital and labor. The distribution of the county fixed effects is shown in Fig. 5. The figure illustrates that there is substantial heterogeneity in manufacturing productivity across US counties. The county at the top of the distribution is 2.9 times more productive than the county at the bottom of the distribution. Log TFP in the counties at the 10th percentile, median, and 90th percentile are 1.54, 1.70 and 2.20, respectively.

Figure 6 shows how TFP has changed over time. Specifically, it plots average TFP by county in 1977 on the x-axis against average TFP by county in 1992 on the y-axis.[6]

[5] Regressions are weighted by plant output. The sample is restricted to counties that had 10 or more plants in either 1977 or 1992 in the 2xxx or 3xxx SIC codes. There are 2126 counties that satisfy the sample restriction. For confidentiality reasons, any data from counties whose output was too concentrated in a small number of plants are not in the figure (although they are included in the regression).

[6] There are 1951 counties for which data could be released by the Census. TFP estimates for each year come from separate regressions for 1977 and 1992.
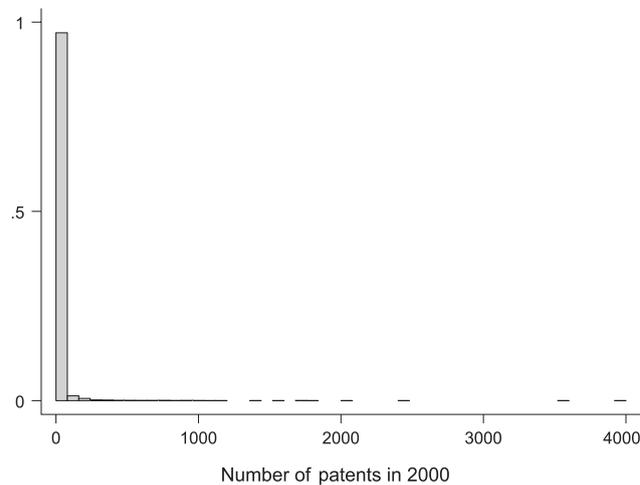
Figure 7 *Distribution in the number of patents filed by city.* Notes: The figure reports the distribution of the average yearly number of patents filed between 1998 and 2002 across cities. I use the average over 5 years to reduce small sample noise. The level of observation is the city, as reported in the patent file. This definition of city does not correspond to the definition of metropolitan statistical area.

The regression line comes from a regression of 1992 TFP on 1977 TFP weighted by the inverse of the county fixed effects' standard errors.[7] The coefficient is 0.919 (0.003), indicating a high degree of persistence of TFP over time. This coefficient is lower than the corresponding coefficient for nominal wages in Fig. 3. This difference may indicate that changes in productivity are not the only driver of changes in nominal wages across locations. Alternatively it may indicate that average productivity is measured with more error than average wages and therefore displays more mean reversion. It is plausible that measured productivity contains more measurement error than measured wages because productivity is inherently more difficult to measure and because the sample of plants available in the Economic Census is smaller than the sample of workers available in the Census of Population.

2.4. Innovation

Innovative activity is even more concentrated than overall economic activity. One measure of innovation is the number of patents filed. Figure 7 shows the distribution of the number of utility patents filed by each city per year from 1998 to 2002.[8] The level of observation here is the city, as reported in the patent file. Unlike in the rest of

[7] This regression does not include an intercept, because both the dependent variable and independent variable come from separate regressions that include separate constants.

[8] I include 5 years instead of one to reduce sample noise. Data are from the NBER Patent Database. Utility patents are typically granted to those who invent or discover a new and useful process or machine.

the paper, in this figure and the next figure the definition of city does not correspond to metropolitan statistical area. The figure shows that most cities generate either no patents or a limited number of patents each year. On the other hand there is a handful of cities that file a very large number of patents. Conditional on generating at least 1 patent in the five years between 1998 and 2002, the median city generates only an average of .4 patents per year, while the city at the 75% percentile has only 2 patents per year. By contrast, the two cities at the top of the distribution—Santa Clara, CA, in the heart of Silicon valley and Armonk, NY, where IBM is located—generate 3390 and 3630 patents respectively. Houston, San Jose and Palo Alto follow with 2399, 1906 and 1682 patents per year, respectively.[9] Overall, it is pretty clear that the creation of new technologies and new products is highly spatially concentrated.

Importantly, there is little evidence that the geographic concentration of innovative activity is diminishing over time. Indeed, the spatial distribution of innovation appears remarkably stable over the last 2 decades. This is shown in Fig. 8, where I plot the average yearly number of patents filed in the 1978-1982 period on the x-axis against the average yearly number of patents filed in the 1998-2002 period on the y-axis. The sample includes all cities with at least 1 patent filed in either period. For visual clarity, the figure excludes 3 cities that have more than 2000 patents per year. The regression coefficient (std error) is 1.009 (0.0311), with intercept at 15.41 (2.26). (The regression and the fitted line in the figure are both based on the full sample that includes the 3 cities with more than 2000 patents.) The regression indicates that there is no evidence of convergence in innovative activity. The number of patents per city has increased in this period, but the increase is exactly proportional to the 1980s level.

3. EQUILIBRIUM IN LOCAL LABOR MARKETS

In the previous Section I have documented large and persistent differences in productivity and wages across local labor markets within the US. In this section, I present a simple general equilibrium framework intended to address two questions. First, how can these differences persist in equilibrium? Second, what are the effects of these differences on workers in different cities?

Ever since the publication of the models by Rosen (1979) and Roback (1982), the Rosen–Roback framework has been the general equilibrium model most frequently used to model shocks to local economies. For this reason, Glaeser (2001) defines the Rosen–Roback framework "the workhorse of spatial equilibrium analysis". The main reasons for its popularity are its simplicity, tractability, and especially the fact that it captures a very intuitive notion of equilibrium across local labor markets within a country. In its most basic and most commonly used version (Roback, 1982, Section I), the model assumes that:

[9] The city at the 99% percentile generates 178 patents.
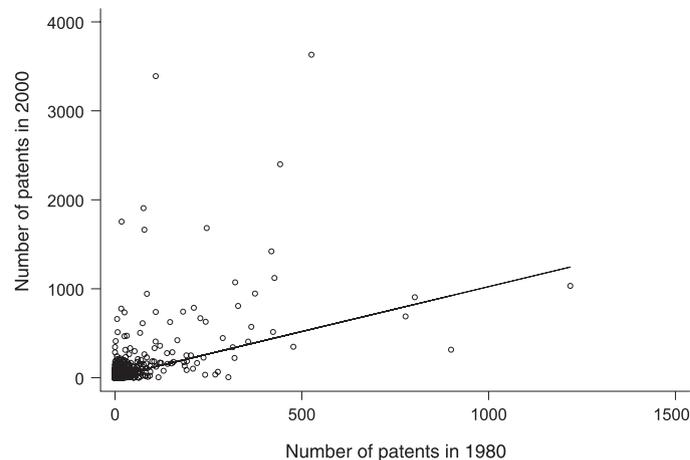
Figure 8   *Change over time in the number of patents filed by city.* Notes: The x-axis is the average yearly number of patents filed between 1978 and 1982. The y-axis is the average yearly number of patents filed between 1998 and 2002. I use the averages over 5 years to reduce sample noise. The level of observation is the city, as reported in the patent file. This definition of city does not correspond to metropolitan statistical area. For visual clarity, the figure excludes 3 cities that have more than 2000 patents per year. A regression based on the full sample (i.e. including the cities with more than 3000 patents per year) yields a coefficient (std error) equal to 1.009 (0.0311). The fitted line in the figure is based on the full sample (i.e. including the 3 cities with more than 2000 patents per year).

1. Each city is a competitive economy that produces a single internationally traded good using labor, land and a local amenity. Technology has constant returns to scale
2. Workers' indirect utility depends on nominal wages, cost of housing and local amenities
3. Labor is homogenous in skills and tastes[10] and each worker provides one unit of labor
4. Labor is perfectly mobile so that the local labor supply is infinitely elastic
5. Land is the only immobile factor and its supply is fixed.

In its simplest form, and the one that is most commonly used in the literature (Roback, 1982, Section I), the Rosen–Roback key insight is that any local shock to the demand or supply of labor in a city is, in equilibrium, *fully* capitalized in the price of land. As a consequence, shocks to a local economy do not affect worker welfare. Consider, for example, a productivity shock that makes workers in city c more productive than workers in other cities. In the Rosen–Roback framework, the increase in productivity in city c results in an increase in nominal wages in city c and a similar increase in housing costs in city c, so that in equilibrium workers are completely indifferent between city c and all the other cities. In the new equilibrium, workers are more productive but they are not better

[10] Roback (1988) considers the cases of heterogenous labor.

off. The owners of land in city c are better off, by an amount equal to the productivity increase. This result depends on the assumption that the local labor supply is infinitely elastic and that the elasticity of housing supply is limited.[11]

The assumptions of this model are restrictive, and rule out some interesting questions regarding the incidence of localized shocks to a local economy. In this section, I present a more general equilibrium framework that seeks to take the spatial equilibrium model a step closer to reality. The goal of the model is to clarify what happens to wages, costs of housing and worker utility when a local economy experiences a shock to labor demand or labor supply. An example of a shock to labor demand is an increase in productivity. An example of a shock to labor supply is an increase in amenities. I assume that workers and firms are mobile across cities, but worker mobility is not necessarily infinite, because workers have idiosyncratic preferences for certain locations. Moreover, housing supply is not necessarily fixed. This implies that the elasticity of local labor supply is not necessarily infinite and the elasticity of housing supply is not necessarily zero. In this context, shocks to a local economy are not necessarily fully capitalized into land prices. This is important, because it allows for interesting distributional and welfare implications. The model clarifies exactly how the welfare consequences of localized labor market shocks depend on the relative magnitude of the elasticities of local labor supply and housing supply.

In Section 3.1 I describe the case of homogenous labor. It is a useful and transparent starting point. It clarifies the role that the elasticity of labor and housing supply play in determining how shocks to a local economy affect workers' utility. In reality, however, workers are not all homogenous but they differ along many dimensions, most notably in their skill level. Moreover, shocks to local economies rarely affect all workers equally. Instead, shocks to local economies are often skill-biased in the sense that they shift the demand for some skill groups more than others. For these reasons, in Section 3.2 I describe the more general case of heterogenous labor. In Section 3.3 I allow for agglomeration economies. In Section 3.4 I discuss the case where there are multiple industries within each local economy and local multipliers. In Section 3.5 I review some of the existing empirical evidence.

---

[11] In the simplest form of the model, there is one margin of adjustment that allows to accommodate some in-migration to the more productive city. While land is assumed to be fixed, workers can adjust their consumption of housing. When housing prices increase in city c, each worker consumes a little less housing. This allows a small increase in the number of workers in the more productive city, even with fixed land. In a more general version of the model, Roback (1982, Section II) keeps the assumption that land is fixed but allows for the production of housing. Housing production is assumed to use labor, which is perfectly mobile, and land, which is fixed. In this version of the model, there are two margins of adjustment that allow to accommodate in-migration to a city. First, like before, workers can adjust their consumption of housing in response to increases in housing prices. Second, unlike before, the housing stock can increase in response to increased demand. In this version of the model, more workers change city after a city-specific productivity shock. However, the key implication for incidence of the shock does not change. Because workers are assumed to be perfectly mobile and homogenous, their utility is never affected by the shock.

Over the years, many versions of the spatial equilibrium model have been proposed. The version of model that I present is based on Moretti (forthcoming). The proposed framework is based on assumptions designed to make it very simple and transparent while at the same time not unrealistic. The model seeks to describe spatial equilibrium in the long run and is probably not well suited to describe year to year adjustments.[12] Topel (1986) and Glaeser (2008) propose alternative equilibrium frameworks that take into account the dynamics of wages and employment. Roback (1982), Glaeser (2008) and Glaeser and Gottlieb (2009) propose frameworks where housing production uses both local labor (and of course land). By contrast, in my simplified framework housing production does not use local labor. Combes et al. (2005) link the spatial equilibrium framework to some of the insights from the New Economic Geography literature.

3.1. Spatial equilibrium with homogeneous labor
3.1.1. Assumptions and equilibrium
I begin by considering the case where there is only one type of labor. As in Rosen–Roback, I assume that each city is a competitive economy that produces a single output good y which is traded on the international market, so that its price is the same everywhere and set equal to 1. Workers and firms are mobile and locate where utility and profits are maximized. Like in Roback, I abstract from labor supply decisions and I assume that each worker provides one unit of labor, so that local labor supply is only determined by workers' location decisions. The indirect utility of worker i in city c is

$$U_{ic} = w_c - r_c + A_c + e_{ic} \tag{1}$$

where w_c is the nominal wage in city c; r_c is the cost of housing; A_c is a measure of local amenities.[13] The random term e_{ic} represents worker i idiosyncratic preferences for location c. A larger e_{ic} means that worker i is particularly attached to city c, holding constant real wage and amenities. For example, being born in city c or having family in city c may make city c more attractive to a worker irrespective of city c's real wages and amenities. Assume that there are two cities: city a and city b and that worker i's relative preference for city a over city b is

$$e_{ia} - e_{ib} \sim U[-s, s]. \tag{2}$$

The parameter s characterizes the importance of idiosyncratic preferences for location and therefore the degree of labor mobility. If s is large, it means that preferences for location are important and therefore worker willingness to move to arbitrage away real wage differences or amenity differences is limited. On the other hand, if s is small,

---

[12] The reason is that in the short run, frictions in labor mobility and in housing supply may constrain the ability of workers and housing stock to fully adjust to shocks.

[13] In Roback's terminology, A_c is a consumption amenity.

preferences for location are not very important and therefore workers are more willing to move in response to differences in real wages or amenities. In the extreme, if $s = 0$ there are no idiosyncratic preferences for location and therefore worker mobility is perfect.

While parsimonious, the model captures the four most important factors that might drive worker mobility: wages, the cost of living, amenities, and individual preferences. A worker chooses city a if and only if $e_{ia} - e_{ib} > (w_b - r_b) - (w_a - r_a) + (A_b - A_a)$. In equilibrium, the marginal worker needs to be indifferent between cities. This equilibrium condition implies that local labor supply is upward sloping, and its slope depends on s. For example, local labor supply in city b is

$$w_b = w_a + (r_b - r_a) + (A_a - A_b) + s\frac{(N_b - N_a)}{N} \tag{3}$$

where N_c is the endogenously determined log of number of workers in city c; and $N = N_a + N_b$ is assumed fixed. The key point of Eq. (3) is that the elasticity of local labor supply depends on worker preferences for location. If idiosyncratic preferences for location are very important (s is large), then workers are relatively less mobile and the elasticity of local labor supply is low. In this case, the local labor supply curve is relatively steep. If idiosyncratic preferences for location are not very important (s is small), then workers are relatively more mobile and the elasticity of local labor supply is high. In this case, the local labor supply curve is relatively flat. In the case of perfect mobility ($s = 0$), the elasticity of local labor supply is infinite and the local labor supply curve is perfectly flat. In that case, any difference in real wages or in amenities, however small, results in an infinitely large number of workers willing to leave one city for the other.[14] The intercept in Eq. (3) indicates that, for a given slope, if the real wage in city a increases or local amenities improve, workers leave city b and move to city a.

An important difference between the Rosen–Roback setting and this setting is that in Rosen–Roback, all workers are identical, and always indifferent across locations. In this setting, workers differ in their preferences for location. While the marginal worker is indifferent between locations, here there are inframarginal workers who enjoy economic rents. These rents are larger the smaller the elasticity of local labor supply.[15]

---

[14] Tabuchi and Thisse (2002) model how worker heterogeneity generates an upward sloping local labor supply and how this affects the spatial distribution of economic activity.

[15] It is not easy to obtain credible empirical estimates of the elasticity of local labor supply. First, one needs to isolate labor market shocks that are both localized and demand driven. Second, one needs to identify the effect both on wages and land prices. For example, Greenstone et al. (forthcoming) document that the exogenous opening of a large manufacturing establishment in a county is associated with a significant increase in employment and local nominal wages. The wage increase appears to persist five years after the opening of the new plant. However, this result per se does not necessarily imply that local labor supply is upward sloping. As Eq. (3) indicates, what matters in this respect is whether the demand-driven shift in employment causes wages to increase over and above land costs. The finding that an increase in the local demand for labor results in an increase in local wages does not per se imply that local labor supply is upward sloping. In principle, such finding is consistent with a spatial equilibrium where the local supply of labor is infinitely elastic but the supply of housing is inelastic.

The production function for firms in city c is Cobb–Douglas with constant returns to scale, so that

$$\ln y_c = X_c + hN_c + (1 - h)K_c \qquad (4)$$

where X_c is a city-specific productivity shifter;[16] and K_c is the log of capital. I focus first on the case where X_c is given. Later, I discuss the model with agglomeration economies in which X_c is a function of density of economic activity or human capital. Firms are assumed to be perfectly mobile. If firms are price takers and labor is paid its marginal product, labor demand in city c is

$$w_c = X_c - (1 - h)N_c + (1 - h)K_c + \ln h. \qquad (5)$$

I assume that there is an international capital market, and that capital is infinitely supplied at a given price i.[17] I also assume that each worker consumes one unit of housing. This implies that the (inverse of) the local demand for housing is just a re-arrangement of Eq. (3):

$$r_b = (w_b - w_a) + r_a + (A_b - A_a) - s\frac{(N_b - N_a)}{N}. \qquad (6)$$

To close the model, I assume that the supply of housing is

$$r_c = z + k_c N_c \qquad (7)$$

where the number of housing units in city c is assumed to be equal to the number of workers. The parameter k_c characterizes the elasticity of the supply of housing. I assume that this parameter is exogenously determined by geography and local land regulations. In cities where geography and regulations make it is easy to build new housing, k_c is small. In the extreme case where there are no constraints to building new houses, the supply curve is horizontal, and k_c is zero. In cities where geography and regulations make it difficult to build new housing, k_c is large. In the extreme case where it is impossible to build new houses, the supply curve is vertical, and k_c is infinite. A limitation of Eq. (7) is that it implicitly makes two assumptions that, while helpful in simplifying the model, are not particularly realistic. First, housing production in this model does not involve the use of any local input. Roback (1982) and Glaeser (2008), among others, discuss spatial equilibrium in the case where housing production involves the use of local labor and other local inputs. Second, Eq. (7) ignores the durability of housing. Glaeser (2008) point out that once built, the housing stock does not depreciate quickly and this introduces an

---

[16] In Roback terminology, X_c is a productive amenity.

[17] In equilibrium, the marginal product of capital has to be equal to $X_c - hK_c + hN_c + \ln(1 - h) = \ln i$.

asymmetry between positive and negative demand shocks. In particular, when demand declines, the quantity of housing cannot decline, at least in the short run. The possible implications of this asymmetry are analyzed by Notowidigdo (2010).

Equilibrium in the labor market is obtained by equating Eqs (3) and (5) for each city. Equilibrium in the housing market is obtained by equating Eqs (6) and (7).

In this framework, workers and landowners are separate agents and landowners are assumed to live abroad. While in reality most workers own their residence, keeping workers separate from landowners has the advantage of allowing me to separately identify the welfare consequences of changes in housing values from the welfare consequences of changes in labor income. This is important both for conceptual clarity and for thinking about the different implications of the results for labor and housing policies.[18]

This model differs from the model of local labor markets proposed by Topel (1986) because it ignores dynamics. This model also differs from most of the existing versions of the spatial equilibrium model in that it describes a closed economy with a fixed number of workers, so that shock to a given city affects the other city. For example, an increase in labor demand in city b affects labor supply, wages and prices in city a. By contrast, most existing versions of the spatial equilibrium model assume that local shocks to a city affect local outcomes there, but have a negligible effect on the rest of the national economy because the rest of the economy is large relative to the city. (See for example: Glaeser (2008, 2001) and Notowidigdo (2010)). In this sense most of the existing models are not truly general equilibrium models.[19]

3.1.2. Effect of a labor demand shock on wages and prices

I begin by considering the effect of an increase in labor demand in city b. This demand increase could be due to a localized technological shock that increases the productivity of firms located in city b. Alternatively, it could be due to an improvement in the product demand faced by firms in city b. Later, I consider the effect of an increase in labor supply in city b.

I assume that in period 1, the two cities are identical and in period 2, total factor productivity increases in city b. Specifically, I assume that in period 2, the productivity shifter in b is higher than in period 1: $X_{b2} = X_{b1} + \Delta$, where $\Delta > 0$ represents a

---

[18] On the other hand, this assumption has the disadvantage that it misses some important features of housing and labor markets. When workers are also property owners, a localized increase in housing values in a city implies both an increase in the value of the asset but also an increase in the user cost of housing. The only way for property owners to access the increased value of the asset is to move to a different city.

[19] In the interest of simplicity, the model completely ignores congestions costs. Equilibrium is achieved only because housing costs in a city increase when population increases. In reality, congestion costs (for example: transportation costs) are probably an another important determinant of equilibrium across cities. Allowing for congestion costs would not alter the qualitative predictions of the model, but it would result in smaller predicted increases in housing costs in cities that experience positive productivity shocks. The reason is simple. As a city becomes more productive and its workforce increases, commuting costs increase, thus reducing its relative attractiveness.

positive, localized, unexpected productivity shock.[20] I have added subscripts 1 and 2 to denote periods 1 and 2. The amenities in the two cities are assumed to be identical and to remain unchanged.

Workers are now more productive in city b than a. Attracted by this higher productivity, some workers move from a to b:

$$N_{b2} - N_{b1} = \frac{N}{N(k_a + k_b) + 2s} \Delta \geq 0. \tag{8}$$

The equation indicates that number of movers is larger the elasticity of labor supply (i.e. the smaller is s) and the larger the elasticity of housing supply in city b (i.e. the smaller is k_b). This is not surprising, because a smaller s implies that idiosyncratic preferences for location are less important, and therefore that labor is more mobile in response to real wage differentials. A smaller k_b means that it is easier for city b to add new housing units to accommodate the increased demand generated by the in-migrants.

The nominal wage in city b increases by an amount equal to the productivity increase:

$$w_{b2} - w_{b1} = \Delta. \tag{9}$$

Because of in-migration, the cost of housing in city b needs to increase. The magnitude of the increase is a fraction of Δ and depends on how elastic is housing supply in b relative to a:

$$r_{b2} - r_{b1} = \frac{k_b N}{N(k_a + k_b) + 2s} \Delta \geq 0. \tag{10}$$

This increase in housing costs is larger the smaller the elasticity of housing supply in city b (large k_b) relative to city a. Because nominal wages increase more than housing costs (compare Eqs (9) and (10)), real wages increase in b:

$$(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = \frac{k_a N + 2s}{N(k_a + k_b) + 2s} \Delta \geq 0. \tag{11}$$

Although the original productivity shock only involves city b, in general equilibrium, prices in city a are also affected. In particular, out-migration lowers the cost of housing.[21]

---

[20] I am modeling the productivity shock as an increase in total factor productivity. Results are similar in the case where the shock only increases productivity of labor.

[21] The change in the cost of housing in a is

$$r_{a2} - r_{a1} = -\frac{k_a N}{N(k_a + k_b) + 2s} \Delta \leq 0. \tag{12}$$

Because the nominal wage in a does not change,[22] the net effect is an increase in real wages in a:

$$(w_{a2} - w_{a1}) - (r_{a2} - r_{a1}) = \frac{k_a N}{N(k_a + k_b) + 2s} \Delta \geq 0. \tag{13}$$

It is important to note that in general, real wages differ in the two cities in period 2. In particular, a comparison of Eq. (11) with Eq. (13) indicates that in period 2 real wages are higher in city b. This is not surprising, because city b is the one directly affected by the productivity shock. While labor mobility causes real wages to increase in city a as well, real wages are not fully equalized because mobility is not perfect in that only the marginal worker is indifferent between the two cities in equilibrium. With perfect mobility ($s = 0$), real wages are completely equalized because *all* workers need to be indifferent between the two cities.[23]

The marginal worker in period 2 is different from the marginal worker in period 1. Since city b offers higher real wages in period 2, the new marginal worker in period 2 has stronger preferences for city a. In particular, the change in the relative preference for city a of the marginal worker is equal to[24]

$$(e_{a2} - e_{b2}) - (e_{a1} - e_{b1}) = \frac{2s\Delta}{N(k_a + k_b) + 2s} \geq 0. \tag{14}$$

Note that firms are indifferent between cities. Because of the assumptions on technology, firms have zero profits in both cities. While labor is now more expensive in b, it is also more productive there. Because firms produce a good that is internationally traded, if skilled workers weren't more productive, employers would leave b and relocate to a.

In the production function used here, all firms in a city are assumed to share a city-specific productivity shifter. The implicit assumption is that any city-specific characteristic affects all firms equally. For example, the transportation infrastructure, the weather, local institutions, local regulations, etc. affect the productivity of all producers in the same way. It would be easy to extend this framework to allow for an additional firm-city specific productivity shifter:

$$\ln y_{jc} = (X_c + X_{jc}) + hN_{jc} + (1 - h)K_{jc} \tag{15}$$

---

[22] This may look surprising at first. Given that the number of workers has declined, and that the demand curve is downward sloping, one might expect an increase in wages of those workers who stay in a. Indeed, this would be true in a model without capital. But in a model that includes capital, the amount of capital used by firms declines in b and increases in a. This capital flow off-sets the changes in labor supply.

[23] To see this, compare Eq. (11) with (13), setting $s = 0$.

[24] This change is by construction equal to the change in the difference in real wages between the two cities.

where j indexes a firm, X_c is a productivity effect shared by all firms in city c, and X_{jc} is a productivity effect that is specific to firm j and city c. This formulation allows some firms to benefit more from some city characteristics than others. For example, the specific type of local infrastructure in a given city may affect the TFP of some firms more than others. This is analogous to introducing individual specific location preferences in workers' utility functions. For the same reason that preferences for location make workers less responsive to differences in real wages across locations, the term X_{jc} makes firms less mobile. Effectively, some firms enjoy economic rents generated by their location-firm specific match. Small differences in production costs may not be enough to induce these firms to relocate, in the same way that worker idiosyncratic preferences for location lower the elasticity of labor supply.

3.1.3. Incidence: who benefits from the productivity increase?

In this setting, the benefit of the increase in productivity Δ is split between workers and landowners.[25] Eqs (10)–(13) clarify that the incidence of the shock depends on which of the two factors—labor or land—is supplied more elastically at the local level. In turn, the elasticity of local labor supply and the elasticity of housing supply ultimately depend on the preference parameter s and the supply parameters k_a and k_b. For a given elasticity of housing supply, a lower local elasticity of labor supply implies that a larger fraction of the productivity shock in city b accrues to workers in city b, and a smaller fraction accrues to landowners in city b. Intuitively, when labor is relatively less mobile, it captures more of the economic rent generated by the productivity shock. A lower local elasticity of labor supply also implies a smaller increase in real wages in the non affected city (city a), since the channel that generates benefits for the non affected city is the potential for worker mobility.

On the other hand, for a given elasticity of labor supply, a lower elasticity of housing supply in city b relative to city a (k_b bigger than k_a) implies that housing quantity adjusts less in city b following the productivity shock. As a consequence, housing prices increase more and a larger fraction of the productivity gain accrues to landowners in city b and a smaller fraction accrues to workers.

The role played by the elasticity of labor and housing supply in determining the incidence of the productivity shock between workers and landowners and between city a and city b is clearly illustrated in four special cases.

1. If labor is completely immobile ($s = \infty$), Eq. (11) becomes $(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = \Delta$, indicating that real wages in city b increase by the full amount of the productivity shock. In this case, the benefit of the shock accrues entirely to workers in city b. The intuition is that if labor is a fixed factor, workers in the city hit by the shock

---

[25] By construction: Δ = change in real wage in a + change in real wage in b + change in land price in a + change in land price in b.

capture the full economic rent generated by the shock. Nothing happens to workers in a, as their real wage is unchanged: Eq. (13) becomes $(w_{a2} - w_{a1}) - (r_{a2} - r_{a1}) = 0$. Moreover, since no worker moves in equilibrium, housing prices in both cities remain unchanged so that landowners are indifferent. For example, Eq. (10) becomes $r_{b2} - r_{b1} = 0$, indicating that housing prices in b are not affected.

2. If labor is perfectly mobile ($s = 0$), Eqs (11) and (13) become: $(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = (w_{a2} - w_{a1}) - (r_{a2} - r_{a1}) = \frac{k_a}{k_a+k_b}\Delta$. Because of perfect labor mobility, real wages need to be identical in a and b, otherwise workers would leave one city for the other. In this case, incidence depends on the relative elasticities of housing supply in the two cities. To see this, note that the increase in real wages is a fraction $\frac{k_a}{k_a+k_b}$ of Δ. The rest of Δ accrues to landowners in b, since housing prices in b increase by $r_{b2} - r_{b1} = \frac{k_b}{k_a+k_b}\Delta$. The fraction that accrues to workers depends on which of the two cities has more elastic housing supply. For example, if the elasticity of housing supply is the same in a and b, than we have an equal split between workers and landowners, with real wages in both cities increasing by $\frac{1}{2}\Delta$, and land prices in b increasing by $\frac{1}{2}\Delta$. On the other hand, if the elasticity of housing supply is larger in city b then landowners capture less of the total economic rent, because their factor is more elastically supplied in the city originally hit by the shock.

3. If housing supply in b is fixed ($k_b = \infty$), the entire productivity increase is capitalized in land values in city b. This is the Rosen–Roback case described above. City b becomes more productive but it cannot expand its workforce because housing cannot expand. No one can move to city b, and the only effect of the productivity shock is to raise cost of housing by $r_{b2} - r_{b1} = \Delta$. All the benefit goes to landowners in b. Real wages are not affected, and workers in both cities are indifferent. This is a case where, even in the presence of a shock that makes some firms more productive, labor is prevented from accessing this increased productivity by the constraints on housing supply. Part of the increase in productivity is therefore wasted.

4. If housing supply in b is infinitely elastic ($k_b = 0$), then Eq. (10) becomes $r_{b2} - r_{b1} = 0$, indicating that housing prices in b do not change. For each additional worker who intends to move to city b, a housing unit is added so that housing prices never increase. Landowners are indifferent, and the entire benefit of the productivity increase accrues to workers. Equation (11) becomes $(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = \Delta$, indicating that real wages in city b increase by the full amount of the productivity shock. Real wages in city a also increase, but less than in b: $(w_{a2} - w_{a1}) - (r_{a2} - r_{a1}) = \frac{k_a N}{N k_a + 2s}\Delta$.

3.1.4. Effect of a labor supply shock on wages and prices

So far, I have focused on what happens to a local economy following a shock generated by a labor demand shift. What distinguishes city b from city a, is that in city b the demand for labor is higher. I now discuss the opposite case, where a local economy experiences an increase in the supply of labor. Specifically, I consider what happens when city b becomes

more desirable for workers relative to city a. I assume that in period 2, the amenity level increases in city b: $A_{b2} = A_{b1} + \Delta'$, where $\Delta' > 0$ represents the improvement in the amenity. I assume that the amenity level in a does not change, and that productivity is the same in the two cities.[26]

As in the case of a demand shift above, $\frac{N}{N(k_a+k_b)+2s}\Delta'$ workers move from a to b. As before, the cost of housing increases in b (by the amount in Eq. (10)) and declines in a (by the amount in Eq. (12)). Also, similar to before, the nominal wage in a does not change. A difference with the demand shock case is that the nominal wage in b does not increase, but it remains unchanged.[27]

As a consequence, real wages decline in city b

$$(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = -\frac{k_b N}{N(k_a + k_b) + 2s}\Delta' \leq 0 \tag{16}$$

and increase in city a:

$$(w_{a2} - w_{a1}) - (r_{a2} - r_{a1}) = \frac{k_a N}{N(k_a + k_b) + 2s}\Delta' \geq 0. \tag{17}$$

Intuitively, workers are willing to take a negative compensating differential in the form of lower real wages to live in the more desirable city. Landowners in b experience an increase in their property values, while landowners in a experience a decline.

The incidence of the shock is similar to what I discuss in Section 3.1.3. As with the case of a demand shock, the exact magnitude of workers' and landowners' gains and losses depend on the elasticity of labor supply and the elasticity of housing supply. The four special cases outlined in Section 3.1.3 apply to this case as well.

3.2. Spatial equilibrium with heterogenous labor

In Section 3.1, I have considered the case where all workers are identical in terms of productivity. In this section, I consider the case where there are 2 types of workers: skilled workers (type H) and unskilled workers (type L). I assume that skilled and unskilled workers in the same city face the same housing market. I discuss what happens in

---

[26] Here, the labor supply increase is a consequence of an increase in amenities, holding constant tastes. Results are similar if one assumes that amenities are fixed, but the taste for those amenities increases.

[27] This may seem counterintuitive at first. One might expect wage *decreases* in response to supply increases. Why do nominal wages not decline in b after it has become more attractive? After all, workers should be willing to take a negative compensating differential in the form of lower nominal wages to live in the more desirable city. Indeed, this is what a model without capital would predict. However, such a model ignores the endogenous reaction of capital. In a model with capital, nominal wages do not move in city b because capital flows to b, offsetting the changes in labor supply. The amount of capital increases in b by $K_{b2} - K_{b1} = \frac{N\Delta'}{N(k_a+k_b)+2s} \geq 0$ and decreases in a by $K_{a2} - K_{a1} = -\frac{N\Delta'}{N(k_a+k_b)+2s} \leq 0$.

equilibrium when the demand for one group changes in one city, while the demand for the other group remains unchanged.

3.2.1. Assumptions and equilibrium

The indirect utilities of skilled workers and unskilled workers in city c are assumed to be, respectively

$$U_{Hic} = w_{Hc} - r_c + A_{Hc} + e_{Hic} \tag{18}$$

and

$$U_{Lic} = w_{Lc} - r_c + A_{Lc} + e_{Lic}. \tag{19}$$

In Eqs (18) and (19), skilled and unskilled workers in a city face the same price of housing so that a shock to the labor demand of one group may be transmitted to the other group through its effect on housing prices.[28] While they have access to the same local amenities, different skill groups do not need to value these amenities equally: A_{Hc} and A_{Lc} represent the skill–specific value of local amenities. Tastes for location can vary by skill group. Specifically, I assume that skilled workers' and unskilled workers' relative preferences for city a over city b are, respectively

$$e_{Hia} - e_{Hib} \sim U[-s_H, s_H] \tag{20}$$

and

$$e_{Lia} - e_{Lib} \sim U[-s_L, s_L]. \tag{21}$$

For example, the case in which skilled workers are more mobile than unskilled workers can be modeled by assuming that $s_H < s_L$.

For simplicity, I focus on the case where skilled and unskilled workers in the same city work in different firms. This amounts to assuming away imperfect substitution between skilled and unskilled workers. This assumption simplifies the analysis, and it is not crucial. The production function for firms in city c that use skilled labor is Cobb–Douglas with constant returns to scale: $\ln y_{Hc} = X_{Hc} + h N_{Hc} + (1-h) K_{Hc}$, where K_{Hc} is the log of capital and X_{Hc} is a skill and city-specific productivity shifter. Similarly, the production

---

[28] It is easy to relax this assumption by assuming residential segregation along skill lines within a city. However, this assumption would not be particularly realistic. Although skilled and unskilled individuals may reside in different parts of a metropolitan area, there always is some overlap which ensures that shocks to a part of the metropolitan area get transmitted to the rest of the area. Empirically, changes in housing prices across neighborhoods within a city are highly correlated.

function for firms that use unskilled labor is $\ln y_{Lc} = X_{Lc} + hN_{Lc} + (1 - h)K_{Lc}$. The rest of the assumptions remain unchanged.[29]

3.2.2. Effect of a labor demand shock on wages and prices

Consider the case where the relative demand for skilled labor increases in b. See Moretti (forthcoming) for the specular case where the relative supply for skilled labor increases in b.

Assume that the productivity of skilled workers increases relative to the productivity of unskilled workers in city b because the productivity shifter for skilled workers in city b is higher in period 2 than in period 1: $X_{Hb2} = X_{Hb1} + \Delta$, where $\Delta > 0$ represents a positive, localized, skill-biased productivity shock. Nothing happens to the productivity of unskilled workers in b and the productivity of skilled and unskilled workers in a. The amenities in the two cities are identical and fixed.

Attracted by higher labor demand, some skilled workers move to b from a. In particular, the number of skilled workers in b increases by

$$N_{Hb2} - N_{Hb1} = \frac{\Delta N((k_a + k_b)N + 2s_L)}{2h(k_a N(s_H + s_L) + k_b N(s_H + s_L) + 2s_H s_L)} \geq 0. \qquad (23)$$

This number depends positively on the elasticity of labor supply for skilled and unskilled workers:

$$\frac{\partial(N_{Hb2} - N_{Hb1})}{\partial s_H} = -\frac{N(k_a N + 2s_L + k_b N)^2 \Delta}{2h(k_a N s_H + 2s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)^2} \leq 0$$

$$(24)$$

and

$$\frac{\partial(N_{Hb2} - N_{Hb1})}{\partial s_L} = -\frac{N^3 \Delta(2k_a k_b + k_a^2 + k_b^2)}{2h(k_a N s_H + 2s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)^2} \leq 0.$$

$$(25)$$

The intuition for the first derivative is obvious: a higher elasticity of labor supply for skilled workers implies that skilled workers are more mobile. The intuition for the second

---

[29] Because skilled and unskilled workers face the same housing market within a city, to obtain the (inverse of) the aggregate demand curve for housing in a city one needs to sum the demand of skilled workers and the demand of unskilled workers. For example,in city b:

$$r_b = \frac{(2s_H s_L)}{(s_H + s_L)} - \frac{(2s_H s_L)(N_{Hb} + N_{Lb})}{N(s_H + s_L)} - \frac{s_L(w_{Ha} - w_{Hb} - r_a)}{(s_L + s_H)} - \frac{s_H(w_{La} - w_{Lb} - r_a)}{(s_L + s_H)}. \qquad (22)$$

derivative is less obvious. A higher elasticity of labor supply for unskilled workers implies that a larger number of unskilled workers move out in response to the inflow of skilled workers, so that the increase in housing costs is more limited which ultimately increases the number of skilled in-migrants.

The number of movers in Eq. (23) also depends positively on the elasticity of housing supply in b:

$$\frac{\partial(N_{Hb2} - N_{Hb1})}{\partial k_b} = -\frac{N^2 \Delta s_L^2}{h(k_a N s_H + 2 s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)^2} \leq 0. \tag{26}$$

A higher elasticity of housing supply (lower k_b) implies that more housing units become available for the incoming skilled workers.

Because skilled workers in b have become more productive, their nominal wage increases by an amount Δ/h, proportional to the productivity increase. Following the inflow of skilled workers, the cost of housing in b increases and the increase is larger the smaller is s_H and the larger is k_b:[30]

$$r_{b2} - r_{b1} = \frac{s_L N k_b \Delta}{h(k_a N s_H + 2 s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)} \geq 0. \tag{27}$$

Skilled workers in both cities experience increases in real wages. In b, the increase in real wages is smaller than the increase in nominal wages because of the increase in the cost of housing:

$$
\begin{aligned}
(w_{Hb2} &- r_{b2}) - (w_{Hb1} - r_{b1}) \\
&= \frac{k_a N s_H + k_b N s_H + k_a N s_L + 2 s_H s_L}{h(k_a N s_H + 2 s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)} \Delta \geq 0.
\end{aligned} \tag{28}
$$

It is easy to see that this change is less than the increase in nominal wages, Δ/h. Since nominal wages don't change and housing costs decline, real wages for skilled workers in a also increase, but by less than in b:

$$
\begin{aligned}
(w_{Ha2} &- r_{a2}) - (w_{Ha1} - r_{a1}) \\
&= \frac{s_L k_a N}{h(k_a N s_H + 2 s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)} \Delta \geq 0.
\end{aligned} \tag{29}
$$

By comparing Eq. (28) with (29), it is easy to confirm that $(w_{Hb2} - r_{b2}) - (w_{Hb1} - r_{b1}) \geq (w_{Ha2} - r_{a2}) - (w_{Ha1} - r_{a1})$.

---

[30] Because of the decline in the number of workers, the cost of housing in a declines by the same amount.

What happens to unskilled workers? In city b their productivity and nominal wages don't change, but housing costs increase. As a consequence, their real wage in b decreases by

$$(w_{Lb2} - r_{b2}) - (w_{Lb1} - r_{b1})$$
$$= -\frac{s_L N k_b}{h(k_a N s_H + 2 s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)} \Delta \leq 0. \quad (30)$$

Effectively, unskilled workers in b compete for scarce housing with skilled workers, and the inflow of new skilled workers hurts unskilled workers through higher housing costs. (For the same reason, the real wage of unskilled workers in a increases.) Since their real wage has declined, the number of unskilled workers in b declines by

$$N_{Lb2} - N_{Lb1} = -\frac{N^2(k_a + k_b)}{2h(k_a N s_H + 2 s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)} \Delta \leq 0. \quad (31)$$

The overall population of city b increases. This is because the number of skilled workers who move to city b is larger than the number of unskilled workers who leave city b. On net

$$(N_{Hb2} + N_{Lb2}) - (N_{Hb1} + N_{Lb1})$$
$$= \frac{\Delta N s_L}{h(k_a N(s_H + s_L) + k_b N(s_H + s_L) + 2 s_H s_L)} \geq 0. \quad (32)$$

An assumption of this model is that skilled and unskilled workers are employed by different firms, so that the labor market is segregated by skill within a city. This assumption effectively rules out imperfect substitutability between skilled and unskilled labor. In a more general setting, skilled and unskilled workers work in the same firm. Most of the results in this section generalize, but the equilibrium depends on the degree of imperfect substitution between skilled and unskilled labor.[31] Specifically, complementarity between skilled and unskilled workers implies that the marginal product of unskilled workers increases in the number of skilled workers in the same firm. Thus, the inflow of skilled workers in city b caused by the increase in their productivity endogenously raises the productivity of unskilled workers in city b. As a consequence, the real wage of unskilled workers declines less than in the case described above. This mitigates the negative effect on the welfare of unskilled workers in city b and it reduces the number of unskilled workers who leave the city.

---

[31] Given that the focus is on skill-biased productivity shocks, a CES technology is more appropriate for the case of integrated labor markets than a Cobb–Douglas technology.

3.2.3. Incidence: changes in wage and utility inequality

The model yields three conclusions regarding the incidence of the skill-biased localized shock.

First, to the extent that mobility is not perfect, a non-degenerate equilibrium is possible. After a shock that makes one group more productive, both groups are still represented in both cities. This conclusion hinges upon the assumption of a less than infinite elasticity of local labor supply. When the productivity shock attracts skilled workers to city b, housing prices increase there, and unskilled workers begin to leave, since their real wage is lower than in city a. The inflow of skilled workers to the city effectively displaces some unskilled workers. In the absence of individual preferences for location, no unskilled worker would remain in city b and the equilibrium would be characterized by complete geographic segregation of workers by skill level. This is clearly not realistic, since in reality we never observe cities that are populated by workers of only one type. In the presence of individual preferences for location, those unskilled workers who have a strong preference for city b over city a opt to stay in city b, even if their real wage is lower in b. Those who leave are those who are less attached to city b.

Therefore, the marginal unskilled worker has weaker preferences for city a after the shock than before the shock. The change in the relative preference for city a of the marginal unskilled worker is equal to

$$
\begin{aligned}
&(e_{La2} - e_{Lb2}) - (e_{La1} - e_{Lb1}) \\
&= -\frac{s_L N(k_a + k_b)}{h(k_a N(s_H + s_L) + k_b N(s_H + s_L) + 2s_H s_L)} \Delta \leq 0.
\end{aligned} \tag{33}
$$

The opposite is true for skilled workers. Because their real wage has increased in city b more than in city a, the marginal skilled worker has stronger preferences for city a after the shock:

$$
\begin{aligned}
&(e_{Ha2} - e_{Hb2}) - (e_{Ha1} - e_{Hb1}) \\
&= \frac{s_H(k_a N + 2s_L + k_b N)}{h(k_a N(s_H + s_L) + k_b N(s_H + s_L) + 2s_H s_L)} \Delta \geq 0.
\end{aligned} \tag{34}
$$

Second, skilled workers in both cities and landowners in city b benefit from the productivity increase. Inframarginal unskilled workers in city b are negatively affected, and inframarginal unskilled workers in city a are positively affected. It is important to highlight that, although inframarginal unskilled workers in city b are made worse off by the decline in their real wage, they are still better off in city b than in city a because of their idiosyncratic preferences for location.

The magnitude of these changes in utility for skilled and unskilled workers and for landowners crucially depends on the elasticities of labor supply of the two groups (which are governed by the preference parameters s_H and s_L) and the elasticities of housing

supply in the two cities (which are governed by the parameters k_a and k_b). The intuition is related to the intuition provided above for the incidence in the case of homogenous labor, although it is complicated by the fact that each group's location decisions affect the other group's utility through changes in housing prices. For example, the gain in real wages experienced in equilibrium by skilled workers in city b is large if their mobility is low (s_H is large):

$$\frac{\partial((w_{Hb2} - r_{Hb2}) - (w_{Hb1} - r_{Hb1}))}{\partial s_H}$$
$$= \frac{(k_a N + 2s_L + k_b N)s_L N k_b \Delta}{h(k_a N s_H + 2s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)^2} \geq 0. \qquad (35)$$

Low mobility implies that fewer skilled workers are willing to leave a and move to b, so that residents of b experience a smaller increase in the cost of housing. Similarly, the gain in real wages experienced in equilibrium by skilled workers in city b is large if the mobility of unskilled workers is high (s_L is small):

$$\frac{\partial((w_{Hb2} - r_{Hb2}) - (w_{Hb1} - r_{Hb1}))}{\partial s_L}$$
$$= -\frac{k_b N^2 s_H \Delta (k_a + k_b)}{h(k_a N s_H + 2s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)^2} \leq 0. \qquad (36)$$

If unskilled workers are highly mobile, more of them leave the city in response to the increase in housing costs. The ultimate equilibrium increase in housing costs is therefore smaller, and this results in a higher real wage (and higher utility) for inframarginal skilled workers in b.

Additionally, the increase in real wages experienced by skilled workers in city b is large if the elasticity of housing supply in b is high (k_b is small), because a high elasticity of housing supply in b implies that for a given increase in city size, the increase in housing costs is small, and this translates into a larger increase into equilibrium real wages for skilled workers:

$$\frac{\partial((w_{Hb2} - r_{Hb2}) - (w_{Hb1} - r_{Hb1}))}{\partial k_b}$$
$$= -\frac{N \Delta s_L (2s_H s_L + k_a N s_H + k_a N s_L)}{h(k_a N s_H + 2s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)^2} \leq 0. \qquad (37)$$

The opposite argument applies to unskilled workers. The decline in their equilibrium real wage in city b depends positively on their elasticity of labor supply:

$$
\frac{\partial((w_{Lb2} - r_{Lb2}) - (w_{Lb1} - r_{Lb1}))}{\partial s_L}
$$
$$
= -\frac{k_b N^2 s_H \Delta (k_a + k_b)}{h(k_a N s_H + 2 s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)^2} \leq 0 \qquad (38)
$$

and negatively on the elasticity of labor supply for skilled workers:

$$
\frac{\partial((w_{Lb2} - r_{Lb2}) - (w_{Lb1} - r_{Lb1}))}{\partial s_H}
$$
$$
= \frac{(k_a N + 2 s_L + k_b N) s_L N k_b \Delta}{h(k_a N s_H + 2 s_H s_L + k_a N s_L + k_b N s_H + k_b N s_L)^2} \geq 0. \qquad (39)
$$

A small elasticity of labor supply for unskilled workers implies that unskilled workers have strong idiosyncratic preferences for location, so that few move in response to the loss in real wage. With perfect mobility ($s_L = 0$), they experience no loss in real wage (See Eq. (30)). Additionally, a large s_H implies that skilled workers have low mobility so that few move in response to the increase in their wage. The ultimate increase in the price of land is therefore small, so the utility loss for inframarginal unskilled workers in b is more contained. With no mobility of skilled workers ($s_H = \infty$), unskilled workers experience no change in the real wages.

For landowners, a higher elasticity of housing supply in city b relative to city a (k_b smaller than k_a) implies that housing quantity adjusts more in city b so that a smaller fraction of the productivity gain accrues to landowners.

A third conclusion of the model is that the difference in nominal wages between skilled and unskilled workers increases nationwide more than the difference in utility between skilled and unskilled workers. To see this, note that the difference between the change in the skilled–unskilled nominal wage gap and the change in the skilled–unskilled utility gap is

$$
\frac{N k D^2 s_L (s_L + 2kN)}{2h^2 (kN s_H + s_H s_L + kN s_L)^2} \geq 0 \qquad (40)
$$

which is non–negative, indicating that the relative nominal wage of skilled workers grows more than their relative utility. The intuition is that the benefits of a higher nominal wage for skilled workers are in part eroded by the higher cost of housing they are exposed to, so that their relative utility does not increase as much as one might have thought just based on the increase in their relative nominal wage (Moretti, forthcoming).

3.3. Spatial equilibrium with agglomeration economies

In Sections 3.1 and 3.2, the productivity of firms in the two cities is determined by the city-specific productivity parameter X_c, which is taken as given. I now consider the case where there are agglomeration economies so that the productivity of firms in a locality is an endogenous function of the level of economic activity in that locality. This amounts to endogenizing the city-specific productivity shifter. For example, one can assume that productivity in a locality is a function of the number of workers in that locality, so that $X_c = f(N_c)$ with $f' > 0$. In this case, the location decisions of workers generates a positive externality. In Section 4 I discuss in detail the possible sources of agglomeration economies.

As in Sections 3.1 and 3.2, a locality that is for some reason more productive attracts more workers. But unlike Sections 3.1 and 3.2, the increase in population has the additional effect of further increasing productivity of local firms. This in turn attracts even more workers and the process continues to the point where land prices are high enough that marginal workers and firms are made indifferent between locations. Most of the results on incidence presented above remain true. The main difference with the previous analysis is that the existence of agglomeration economies has the potential to generate multiple equilibria, with some equilibria characterized by low economic activity, low cost of housing and low nominal wages, and other equilibria characterized by high economic activity, high cost of housing and high nominal wages.[32]

For concreteness, consider the case of homogenous labor and assume a specific functional form for the agglomeration externality: $X_c = x_c + \gamma N_c$, where the parameter γ governs the strength of agglomeration economies. In the version of the model without agglomeration spillovers ($\gamma = 0$), labor demand has the standard downward sloping shape (see Eq. (5)). With agglomeration spillovers, this is not necessarily the case. Eq. (5) becomes

$$w_c = x_c + (\gamma - (1 - h))N_c + (1 - h)K_c + \ln h. \tag{41}$$

An increase in the number of workers employed in a city has two opposing effects. On the one hand, because of the standard assumptions on technology, an increase in the number of workers lowers the marginal product of labor. On the other hand, the increase in population raises labor productivity. If the agglomeration spillover is strong enough ($\gamma > (1 - h)$), the labor demand function in a city may be upward sloping.

As in Section 3.1, assume that the two cities are identical in period 1, and that in period 2 city b experiences an exogenous increase in productivity of size Δ, so that

---

[32] Glaeser (2008) proposes a comprehensive theoretical equilibrium framework with agglomeration externalities. See also Combes et al. (2005) for a useful big-picture graphical treatment of spatial equilibrium with agglomeration economies.

$x_{b2} = x_{b1} + \Delta$. This initial increase in productivity pushes nominal wages up and higher nominal wages attract more workers to city b. The arrival of new workers in city b generates productivity spillovers and, as a consequence, the initial productivity difference is magnified.

It is informative to compare the equilibrium in the case where there are agglomeration spillovers ($\gamma > 0$) with the case where there are no spillovers ($\gamma = 0$). In the presence of spillovers, a productivity shock of size Δ in city b results in an increase in the equilibrium nominal wage that is larger than Δ:

$$w_{b2} - w_{b1} = \frac{h(N(k_a + k_b) + 2s) - \gamma N}{h(N(k_a + k_b) + 2s) - 2\gamma N}\Delta \geq \Delta \geq 0. \tag{42}$$

This is to be expected, because the agglomeration spillover magnifies the effect of the productivity shock. By contrast, in the case with no spillovers ($\gamma = 0$), the increase in nominal wage in city b is exactly equal to Δ. (See Eq. (9)). Not surprisingly, the larger is the magnitude of the agglomeration spillover—i.e. the larger the parameter γ—the larger is the ultimate increase in the equilibrium nominal wage in city b:

$$\frac{\partial(w_{b2} - w_{b1})}{\partial\gamma} = \frac{Nh(N(k_a + k_b) + 2s)\Delta}{(2\gamma N - 2hs - k_b Nh - k_a Nh)^2} \geq 0. \tag{43}$$

Exactly as in Section 3.1, the higher productivity in city b attracts more workers there. The number of workers in city b increases by

$$N_{b2} - N_{b1} = \frac{Nh}{h(N(k_a + k_b) + 2s) - 2\gamma N}\Delta \geq 0. \tag{44}$$

Just as in the standard case without agglomeration spillovers, if housing supply is not infinitely elastic, the increase in the population of city b ultimately results in higher housing costs:

$$r_{b2} - r_{b1} = \frac{k_b Nh}{h(N(k_a + k_b) + 2s) - 2\gamma N}\Delta \geq 0. \tag{45}$$

It is obvious from Eqs (44) and (45) that the increase in city size and the consequent increase in housing costs are larger the larger is the spillover (i.e. the large is γ). If the spillover is zero, Eqs (44) and (45) revert to Eqs (8) and (10) in Section 3.1.

Since both nominal wages and housing costs increase in b, the ultimate effect on real wages is ambiguous and depends on whether the increase in nominal wage is larger or smaller than the increase in housing costs. In particular, the change in the equilibrium

real wage is

$$(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = \frac{(k_a N + 2s)h - \gamma N}{h(N(k_a + k_b) + 2s) - 2\gamma N}\Delta \tag{46}$$

which is clearly smaller than the increase in nominal wage in Eq. (42). This equation indicates that the change in the real wage depends on the magnitude of the spillover relative to other parameters. To see exactly how the change in the equilibrium real wage depends on γ, note that

$$\frac{\partial((w_{b2} - w_{b1}) - (r_{b2} - r_{b1}))}{\partial \gamma} = \frac{Nh(N(k_a - k_b) + 2s)\Delta}{(2\gamma N - 2hs - k_b Nh - k_a Nh)^2} \tag{47}$$

which can be either positive or negative depending on whether $(N(k_a - k_b) + 2s) > 0$ or $(N(k_a - k_b) + 2s) < 0$. If the elasticity of housing supply in city b is larger or equal to the elasticity of housing supply in city a, the derivative is positive, indicating that the change in the equilibrium real wage in city b is positively associated with the strength of the agglomeration spillover γ. In this case, the increase in real wages in Eq. (46) for the case of positive agglomeration spillovers is larger than the increase in real wages in Eq. (11) for the case with no spillovers.

On the other hand, if the elasticity of housing supply in b is small enough relative to the elasticity of housing supply a, the derivative is negative, and the change in the equilibrium real wage in city b is negatively associated with the strength of the spillover.[33] Intuitively, if the elasticity of housing supply in b is small (large k_b), housing prices in b increase more following the productivity shock, and this increase lowers the equilibrium real wages for a given increase in the nominal wage. In the extreme, if the elasticity of housing supply in b is zero ($k_b = \infty$), the equilibrium real wage does not change. To see why, note that if the elasticity of housing supply in b is zero nobody can move to city b because no new housing unit can be added and Eqs (42) and (45) become

$$w_{b2} - w_{b1} = \Delta \tag{50}$$

---

[33] The change in real wages in city a is smaller than the change in b:

$$(w_{a2} - w_{a1}) - (r_{a2} - r_{a1}) = \frac{k_a Nh - \gamma N}{h(N(k_a + k_b) + 2s) - 2\gamma N}\Delta \geq 0. \tag{48}$$

The derivative of this change with respect to γ is

$$\frac{\partial((w_{a2} - w_{a1}) - (r_{a2} - r_{a1}))}{\partial \gamma} = \frac{Nh(N(k_a - k_b) - 2s)\Delta}{(2\gamma N - 2hs - k_b Nh - k_a Nh)^2} \geq 0. \tag{49}$$

and

$$r_{b2} - r_{b1} = \Delta. \tag{51}$$

In this case, the increase in the nominal wage is exactly equal to the productivity shock Δ even in the presence of agglomeration spillovers because the constraint on labor mobility effectively rules out endogenous changes in total factor productivity X_c. The increase in housing prices is exactly equal to the productivity shock Δ because the lack of any response in the supply of housing implies that the entire productivity shock gets capitalized into land prices. Since both nominal wages and housing cost increase by the same amount Δ, Eq. (46) becomes

$$(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = 0. \tag{52}$$

In sum, even with agglomeration economies it is possible (although not necessary) to have a non degenerate equilibrium where both cities have positive population. Qualitatively, the incidence of the productivity shock is not very different from the case discussed above where there are no agglomeration economies.

3.4. Spatial equilibrium with tradable and non-tradable industries

In the model presented above, the only consumption good is a homogenous tradable good. In reality, however, cities produce and consume a variety of goods, both tradable and non–tradable. Here I discuss how this may affect the equilibrium. This discussion is largely based on Moretti (forthcoming).

Assume that there are K tradable industries producing goods $x_1, x_2, x_3, \ldots, x_K$ and M non-tradable industries, producing goods $z_1, z_2, \ldots, z_M$. Consider the case of a positive shock to productivity in tradable industry x_1 in city c. The direct effect of this shock is an increase in employment in industry x_1. The indirect effect is a change in employment both in the rest of the tradable sector and in the non-tradable sector.

Consider first the effect on the non-tradable sector. Following the shock to sector x_1, aggregate income in the city increases for two reasons. First, there are more local jobs; second, if local labor supply is upward sloping, as in Section 3.1, local wages are also higher. The increase in the city budget constraint results in an increase in the local demand for non-tradables $z_1, z_2, \ldots, z_M$. Employment in industries like restaurants, theaters, real estate, cleaning services, legal services, construction, medical services, retail, personal services, etc. grows both because the city has more workers and wages are higher. The magnitude of the multiplier effect depends on three factors. First, it depends on consumer preferences for tradables and non-tradables and the technology in the non-tradable sector. If preferences are such that a larger share of income is spent on locally produced non-tradables, the multiplier is larger, everything else constant. Similarly, a more labor intensive technology in the non-tradable sector results in a larger multiplier,

everything else constant. Second, it depends on the type of new jobs in the tradable sector. Adding skilled jobs in x_1 should have a larger multiplier than adding unskilled jobs, because skilled jobs pay higher earnings and therefore generate a larger increase in the demand for local services.

Third, there are offsetting general equilibrium effects on wages and prices. As explained in Sections 3.1 and 3.2, the magnitude of these effects ultimately depend on the elasticities of local labor and housing supply. If those elasticities are not infinite, local wages, land prices and the price on non-tradables increase following the shock to x_1. In turn, this city-wide increase in labor and land costs causes a decline in the supply of local services. This decline partially—but not fully—undoes the effect of the increase in demand for local services. Effectively, the addition of jobs in x_1 partially crowds out jobs in other industries. If labor supply is locally very elastic, this crowding out is more limited and the increase in labor costs is small, making the multiplier larger.

The shock to industry x_1 may also affect employment in tradable industries $x_2, x_3, \ldots, x_K$, although the direction of the effect is unclear a priori. This effect is governed by three different forces. First, and most importantly, the city-wide increase in labor costs hurts employment in $x_2, x_3, \ldots, x_K$. Because these are tradable industries, the increase in production costs lowers their competitiveness. Unlike in the case of non-tradable goods, the price of tradable goods is set on the national market and cannot adjust to local economic conditions. In the long run, some of the production in these industries is likely to be shifted to different cities. Second, the increase in production of x_1 may increase the local demand for intermediate goods and services. In this case, some elements of the vector $x_2, x_3, \ldots, x_K$ are inputs to produce x_1. This effect depends on the geography of the industry supply chain. While many industries are geographically clustered, the magnitude of this effect is likely to be quantitatively limited if the market for $x_2, x_3, \ldots, x_K$ is truly national. Third, if agglomeration economies are important, the increase in production in sector x_1 may result in more local agglomeration (see Section 4).

Carrington (1996), Moretti (forthcoming) and Black et al. (2005) estimate the employment multiplier at the local level. Carrington (1996) focuses on the short-run multiplier generated by the construction of the Trans-Alaskan Pipeline System. He finds evidence that the increase in construction jobs caused by the Trans-Alaskan Pipeline System had significant multipliers for jobs in other parts of the non-tradable sector in Alaska. In contrast, Moretti (forthcoming) focuses on long-run multipliers. He quantifies the long-term change in the number of jobs in a city's tradable and non-tradable sectors generated by an exogenous increase in the number of jobs in the tradable sector, allowing for the endogenous reallocation of factors and adjustment of prices. He finds that for each additional job in manufacturing in a given city, 1.6 jobs are created in the non-tradable sector in the same city. This effect is significantly larger for skilled jobs: adding one additional skilled job in the tradable sector generates 2.5 jobs in local goods and services, while adding one additional unskilled job in the tradable sector generates 1 job

in local good and services. Industry-specific multipliers indicate that high-tech industries have the largest multiplier. Using a different time horizon, and focusing on time-varying localized shocks to the coal mining sector, Black et al. (2005) uncover smaller multipliers. They find that each additional mining job generates 0.17 non-tradable jobs. Interestingly, the estimated effect is not symmetric. The loss of a mining job results in the loss of 0.34 non-tradable jobs.

Theory suggests that the local multiplier for the tradable sector should be smaller than the one for the non-tradable sector, and possibly even negative. Consistent with this hypothesis, Carrington (1996), Moretti (forthcoming) and Black et al. (2005) fail to find any significant effect of employment in the tradable sector.

The magnitude of local multipliers is important for regional economic development policies. It should be stressed, however, that the presence of large multipliers is not, in itself, a market failure, and therefore does not necessarily justify government intervention.

3.5. Some empirical evidence

The model presented in this Section appears to be general enough to capture many key features of a realistic spatial equilibrium. To get a better sense of how well the model describes the real world, I now review some of the empirical evidence on the assumptions and the predictions of the model.[34]

One type of labor The evidence in Bartik (1991), Blanchard and Katz (1992), and David et al. (1997) is broadly consistent with the predictions of the model with homogenous labor in Section 3.1. These three papers aggregate all workers into one group and find substantial evidence of large labor flows following permanent labor demand shocks to a local labor market. Using state-level variation in demand conditions, Blanchard and Katz (1992) find that the main mechanism that re-establishes the equilibrium after a demand shock appears to be labor mobility, rather than job creation or job migration. Positive demand shocks are followed by substantial in-migration, while negative demand shocks are followed by substantial out-migration, up to the point that the original equilibrium between demand and supply is restored. In other words, these shocks have permanent effects on the size of labor markets. Blanchard and Katz (1992) estimate that it takes slightly less than a decade for the affected state to return to the initial equilibrium after a localized shock. Bartik (1991) estimates suggest a somewhat slower adjustment. The difference is in part due to the fact that the type of shocks examined by Bartik and Blanchard and Katz is different. Both papers focus on permanent shocks. However, Bartik's estimates come from a model where there is a "once-and-for-all shock

---

[34] Glaeser and Gottlieb (2009) argue that the assumption and the conclusions of the spatial equilibrium model seem generally consistent with most first order facts about the US labor and housing markets. For example, geographic mobility in the US is significant, with more than 40% of households changing addresses every 5 years. Yet, Glaeser and Gottlieb note that "there has been little tendency for people to move to high income areas" even in the presence of large wage disparities across areas. However, they caution that the fact that most amenities are hard to measure makes it inherently difficult to test conclusively whether the US labor market is indeed in a spatial equilibrium.

to local job growth, with subsequent growth unchanged from what it would have been". By contrast, Blanchard and Katz's estimates come from a model where the one-time shock to local job growth is allowed to affect subsequent job growth.

Overall, the findings in Blanchard and Katz (1992) on wages and housing prices are consistent with the version of the spatial equilibrium model where both local labor supply and housing supply are quite elastic in the long run. In particular, following a negative shock, nominal wages decline in the short run, but go back to their original level in the long run. Housing prices track changes in nominal wages, so that the decline in real wages is limited.[35]

Topel (1986) generalizes the spatial equilibrium model to a dynamic setting. Consistent with Bartik (1991) and Blanchard and Katz (1992), Topel finds that positive shocks to labor demand in a local labor market increase nominal wages there. But in addition to the other two papers, Topel also finds that it is not just current shocks that matter for current wages, but also expectations of future shocks. In particular, the expectation of a future demand shock to a local labor market generates in-migration to that market and therefore ultimately results in lower current nominal wages. An implication is that wages respond more to transitory shocks to local labor markets than to permanent shocks.

More than one type of labor Topel (1986) is among the first to posit that mobility costs could be different across skill groups, with low-skilled workers having higher costs and suggest that this difference may affect the welfare incidence of demand shocks. Consistent with a version of the model with heterogenous labor in Section 3.2 where the propensity to move of unskilled workers is different from the propensity to move of skilled workers, Topel finds evidence of a larger incidence of localized labor demand shocks on low-skilled workers than on high-skilled workers. In terms of the model this implies that $s_L > s_H$. Recall that the parameter s characterizes the amount of worker preferences for location. A larger s implies a lower elasticity of local labor supply and therefore lower labor mobility in response to economic shocks.[36]

Bound and Holzer (2000) find similar results. Using data on metropolitan areas, Bound and Holzer separately quantify the effects of location-specific labor demand shocks on the labor market outcomes of skilled and unskilled workers. As in Blanchard and Katz, Bound and Holzer (2000) find that positive (negative) labor demand shocks are followed by labor in-migration (out-migration). However, unskilled workers appear to be less sensitive to possible arbitrage opportunities and therefore less mobile following good and bad shocks. This difference in labor mobility has implications for the incidence of the shock. Because of their stronger preferences for location, unskilled workers

[35] Additionally, David et al. (1997) document that the speed of adjustment depends on the exact source of the demand shock.

[36] Evidence in Machin et al. (2009) and in Malamud and Wozniak (2009) indicates that the difference in mobility rates between educated and less educated workers may be causal.

experience significantly larger declines in nominal and real wages following negative demand shocks than skilled workers.[37] As a consequence, it appears that low skilled workers end up suffering more from localized negative demand shocks than high-skilled workers because they see their real wages fall while high-skilled workers move to better labor markets.

Notowidigdo (2010) proposes an alternative explanation for the difference in mobility and incidence between high and low-skilled workers. He posits that low-skilled workers may be shielded from local labor demand shocks because of declining house prices and public assistance programs. Public assistance programs are naturally tilted towards low-skilled workers. Housing price declines may also benefit low-skilled workers more than high-skilled workers if low-skilled workers have higher expenditure shares on housing. In this case, local labor demand shocks have smaller incidence on low-skilled workers than high skilled workers.

Residents vs movers The spatial equilibrium model described above does a good job of characterizing the incidence of demand or supply shocks to a local labor market, when incidence is defined as the share of the gains or losses that accrues to workers and landowners, or the share that accrues to each city. However, the model is poorly equipped to deal with the question of incidence when incidence is about which workers benefit or lose: migrants or original residents. For simplicity, my assumptions completely rule out any labor supply responses by residents, thus forcing all the employment adjustment to come from mobility in and out of the city. Moreover, the model is a full-employment model where involuntary unemployment is ruled out. In reality, however, it is possible that residents may change the amount of labor that they supply following local demand shocks.

The issue of who—between residents and migrants—ends up getting the new jobs created by a positive labor demand shock is clearly important in the presence of involuntary unemployment. This issue is particularly important when thinking about policies aimed at increasing local employment, like local development policies. Implementing a local development policy that increases employment in an area and benefits only migrants from outside the area is quite different politically from implementing a local development policy that benefits residents. This is particularly true if the development policy is financed by local taxpayer money.

The literature disagrees on this point. On one hand, Renkow (2003, 2006) and Partridge et al. (2009) argue that the primary source of employment increases following

---

[37] Bound and Holzer (2000) estimate that a 10% aggregate decline in labor demand in a city causes the nominal wage of high-school and college graduates to decline in the long run by 7% and 4%, respectively. The difference is even larger for younger workers. They also find declines in real wages for both groups, although smaller than the declines in nominal wages. These findings are qualitatively consistent with the estimates in Blanchard and Katz (1992), but they require an elasticity of local labor supply that is lower than the one implied by the estimates in Blanchard and Katz. Similarly, Topel (1986) finds large wage changes following a localized shock for groups of workers with low mobility, and small wage changes for groups of workers with high mobility.

localized demand shocks comes from non-residents. The evidence in these studies, however, is not particularly convincing and is far from conclusive. A more convincing set of empirical studies is represented by Eberts and Stone (1992) and Bartik (1991, 2001). These studies find significant increases in the labor force participation of residents following localized labor demand shocks. In a authoritative review of the literature, Bartik concludes that probably 25% of the new jobs are filled by increases in the labor force participation of local residents in the long run, with the remaining 75% going to outsiders. In other words, three out of four new jobs "in a region are filled by persons who otherwise would not have lived there".

4. THE DETERMINANTS OF PRODUCTIVITY DIFFERENCES ACROSS LOCAL LABOR MARKETS

In Section 2, I documented the large and persistent differences in labor market outcomes across metropolitan areas in the US. In Section 3, I clarified how those differences can persist in equilibrium in the long run, and who the ultimate beneficiaries of those differences are. In that section, the focus is on the consequences of these differences, while the source of these differences is taken as given. City b is assumed to be more productive than city a for some exogenous reason and this higher productivity ultimately results in more population, higher wages and higher housing costs.

In reality, however, most productivity differences within a country are unlikely to be exogenous. In this section, I discuss what might determine productivity differences across locations within a country. Economists have long hypothesized that the concentration of economic activity may be explained by agglomeration economies of some kind. Agglomeration of economic activity is particularly remarkable for industries that produce tradable goods, because the areas where economic activity is concentrated are typically characterized by high costs of labor and land. The observation that firms that produce tradable goods locate in areas where economic activity is dense and labor and land costs are high is consistent with the notion that those areas enjoy agglomeration advantages that offset the higher production costs.

I review the existing empirical evidence on agglomeration economies, focusing on two related questions:

- What do we know about the magnitude of agglomeration economies? (Section 4.1)
- What do we know about the micro mechanisms that generate agglomeration economies? (Section 4.2)

The past two decades have seen significant amounts of effort devoted to answering these two questions. The key empirical challenge has been the possible existence of unobserved features of localities that affect firm productivity even in the absence of agglomeration economies. The main conclusion of this section is that the existing literature has made some progress in empirically testing for the existence and quantifying

the magnitude of agglomeration economies, accounting for possible omitted variables. However, the channels that generate these economies remain more elusive. Much remains to be done in terms of empirically understanding the exact mechanisms that generate agglomerations of economic activity.

The discussion in this section focuses on papers that might be of particular interest to labor economists. See Duranton and Puga (2004a,b), Rosenthal and Strange (2004) and Glaeser and Gottlieb (2009) for recent surveys that are more focused on urban economics.

4.1. Empirical estimates of agglomeration economies

Two empirical approaches have been proposed to test for and quantify agglomeration economies. The first approach is based on the equilibrium location decisions of firms and seeks to infer the importance of agglomeration forces from the observed geographic distribution of employment. Empirically measuring the degree of agglomeration of different localities is not straightforward. Naive indexes of concentration are sensitive to heterogeneity in firm size and in the size of the geographic areas for which data is available. In a landmark study, Ellison and Glaeser (1997) propose a "dartboard" style methodology for comparing the degree of geographic concentration across industries, accounting for differences in firm size and in the definition of geographic units. They find that firms are spread quite unevenly across localities. Almost all industries appear to be localized to some degree, although in many industries, the degree of localization is not large. When industry is defined at the two digit level, high levels of concentration are observed in the tobacco, textile and leather industries. Low levels of concentration are observed in the paper, rubber and plastics, and fabricated metal products industries. In an important follow up, Ellison et al. (forthcoming) use data from the Longitudinal Research Database to compute pairwise coagglomeration measurements for manufacturing industries and relate these coagglomeration measurements to industry characteristics. They document that coagglomeration rates are higher between industries that are economically similar, suggesting that agglomeration advantages may depend both on physical proximity and on economic linkages between firms. In a related study, Rosenthal and Strange (2003) measure the extent of agglomeration by focusing on the localization decisions of new plants. In the presence of mobility frictions, the localization of new plants is particularly informative because it is arguably less constrained by past localization decisions. The empirical results are consistent with the notion that agglomeration economies decline rapidly over space. Duranton and Overman (2005) propose an alternative measure of agglomeration. Unlike the Ellison and Glaeser (1997) measure, the Duranton and Overman measure is based on a continuous measure of distance and therefore does not depend on arbitrary definitions of geographic units. Using data from the UK, they confirm the existence of a significant amount of spatial agglomeration.[38]

---

[38] In particular, they find that more than half of the 4-digit industries in the UK are characterized by a degree of agglomeration that is statistically significant.

While these dartboard tests are informative in quantifying agglomeration, the main challenge in interpreting these tests is that firms base their location decisions on where their profits are expected to be highest, and this could be due to spillovers, natural advantage, or other unobserved cost shifters. The mere existence of agglomeration is not conclusive evidence of agglomeration economies. A second approach to testing for agglomeration economies directly asks whether productivity is higher in areas that are economically denser. An obvious difficulty is that productivity is an elusive quantity that is hard to measure empirically. In practice, existing studies have used alternative measures of productivity, including output per worker, wages and total factor productivity.

Using data on output per worker, Sveikauskas (1975) and Ciccone and Hall (1996) show that increases in employment density in a location are correlated with significant increases in output per worker, although the lack of a solid identification strategy precludes strong conclusions about causality. In an influential paper, Glaeser and Mare (2001) use wages to measure the marginal product of labor, and ask whether wages are higher in large urban areas. Because they use a longitudinal dataset that follows workers over an extended period of time, and some workers are observed moving in and out of urban areas, the authors are able to account for permanent unobserved worker heterogeneity. Consistent with the existence of agglomeration economies, Glaeser and Mare' find a significant wage premium associated with urban areas. The wage profiles of movers indicate that a significant fraction of the urban wage premium accrues to workers over time and stays with them when they leave cities. This finding is consistent with the notion that urban areas speed the accumulation of human capital and that a significant part of the urban wage premium is due to faster productivity growth in urban areas. On the other hand, using French data, Combes et al. (2009) find evidence of significant sorting of high ability workers into urban areas. Estimates of the relationship between wages and density conditional on worker fixed effects are 50% smaller than the unconditional relationship, indicating that at least half of the wage disparity across French cities can be explained by worker quality.[39]

More recent work seeks to provide more direct evidence on the relationship between agglomeration and productivity by testing whether total factor productivity at the firm level is higher in denser areas. Studies in this group use longitudinal plant-level data to estimate firm-level production functions and test whether changes over time in plant output are systematically associated with changes in the characteristics of the area around the plant, after controlling for changes in inputs. Henderson (2003) and Moretti (2004b) are early adopters of this approach. In particular, Henderson (2003) estimates plant level production functions for machinery and high-tech industries as a function of the density of other plants in the area, both in the same industry and in different industries. Identification is based on the longitudinal nature of the data. He finds that in the high-tech sector the number of establishments in an industry is positively associated with
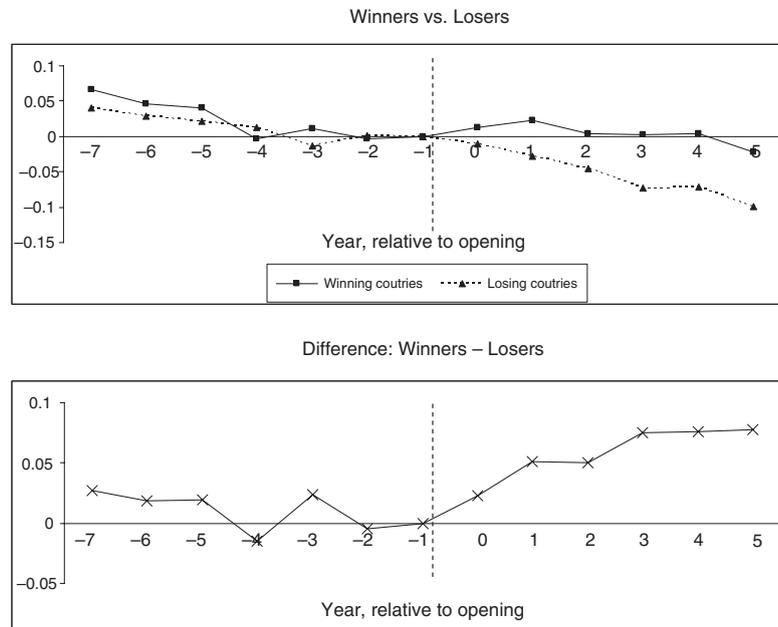
---

[39] See also Wheaton and Lewis (2002).

productivity, although this association is small for the machinery industry. As expected, this association is stronger for plants belonging to single-establishment firms than for plants that belong to multi-establishment firms.

More recently, Greenstone et al. (forthcoming) provide direct estimates of the magnitude of agglomeration economies by comparing the effect of attracting a new manufacturing establishment on the productivity of existing manufacturing establishments in a county. They propose a novel identification strategy that relies not just on the longitudinal nature of the data but also on reported location rankings of profit-maximizing firms to identify a valid counterfactual for what would have happened to the incumbent plants' productivity in the absence of the plant opening. Greenstone, Hornbeck and Moretti find that attracting a new manufacturing establishment to a county results in substantial increases in productivity for existing establishments in that county. Figure 9 replicates their Fig. 1 and shows that five years after the new plant opened, incumbent plants experienced a 12% relative increase in TFP. Consistent with the spatial equilibrium model in Section 3, Greenstone et al. (forthcoming) find that the increased productivity enjoyed by existing plants comes at a cost, as quality-adjusted labor costs increase. As argued in Section 3, this increase is consistent with an upward sloping local supply of labor or an upward sloping supply of housing. Since manufacturing firms produce nationally traded goods and cannot raise output prices in response to higher input prices, the ultimate increase in profits experienced by incumbents is smaller than the productivity increase.

A notable feature of the spatial distribution of economic activity is represented by industry clustering, whereby firms tend to cluster near other "similar" firms (for example: firms that sell similar products). The cluster of IT firms in Silicon Valley, biomedical research in Boston, biotech in San Diego and San Francisco, financial firms in Wall Street and London are notable examples. The findings in Greenstone, Hornbeck and Moretti have interesting implications for explaining the existence and persistence of industrial clusters. The estimated productivity spillover appears to increase with various measures of economic proximity between the new plant and the incumbent plant. Because the documented increase in labor costs applies to all firms in the affected county, while the magnitude of the documented productivity spillovers is larger for pairs of plants that are economically closer, incumbent firms that are economically further away from other firms should become less profitable over time. In the long run, this process may result in increased agglomeration of similar plants in each location.

Arzaghi and Henderson (2008) estimate the effect on the productivity of Manhattan advertising agencies of locating near other advertising agencies. Consistent with the model proposed by Lucas and Rossi Hansberg (2002), they find agglomeration economies characterized by extremely rapid spatial decay. Moreover, consistent with the spatial equilibrium model, the benefit of agglomeration appears to be at least partially offset by higher land prices. In this industry, physical proximity appears to be beneficial because it facilitates networking. Interestingly, the magnitude of the productivity spillover

Winners vs. Losers

Year, relative to opening

Difference: Winners – Losers

Year, relative to opening

Figure 9   *Productivity of incumbent plants in counties with a new plant opening and counterfactual counties, relative to the year of plant opening.* Notes: This figure reproduces Figure 1 in Greenstone et al. (forthcoming). The solid line in the top panel shows average total factor productivity of incumbent plants in counties that successfully attract a new manufacturing plant. In Greenstone, Hornbeck and Moretti these are called winner counties. $t = 0$ represent the year of the new plant opening. The dotted line in the top panel shows average total factor productivity of incumbent plants in counties that bid for the new plant, make it into the group of finalists but ultimately fail to attract the new plant. In Greenstone, Hornbeck and Moretti these are called loser counties. The bottom panel shows the difference in average total factor productivity of incumbent plants between winner counties and loser counties.

appears to vary with firm quality, with higher quality agencies benefitting more from networking than lower quality agencies.

Of course, not every productivity spillover is necessarily a market failure that requires government intervention. Spillovers that occur within a firm, for example, can be in principle internalized. Mas and Moretti (2009) explore how the productivity of a worker varies as a function of the productivity of her co-workers and find evidence of significant within-firm productivity spillovers. The introduction of a high productivity worker in a shift significantly raises the productivity of her co-workers. In particular, substituting a worker with above average permanent productivity for a worker with below average permanent productivity is associated with a 1% increase in the effort of other workers on the same shift. While low productivity workers benefit from the presence of more capable workers, the productivity of high-skilled workers is not hurt by the presence

of low-skilled co-workers. This type of spillover could be internalized by the firm by raising the salary of highly productive workers to reflect their external benefit on the productivity of less productive workers.

While significant progress has been made in estimating agglomeration economies using plausible identifying assumptions, some authors have raised the concern that the observed higher productivity in dense areas may reflect selection due to increased competition. For example, Melitz and Ottaviano (2008) develop a framework where the presence of a larger number of firms in larger markets increases competition and this ultimately causes less efficient firms to disappear. While this argument may apply to countries, it seems less obvious that it should apply to cities within a country. Within a country, firms in the tradable sector compete with firms in other cities, so it seems unlikely that local concentration offers a good measure of the relevant degree of competition. Indeed, Combes et al. (2009) provide recent convincing evidence in support this observation.[40]

4.2. Explanations of agglomeration economies

Understanding the ultimate causes of agglomeration economies is crucial to understanding persistent labor market differences across metropolitan areas. It is also very important for policy, as I discuss in Section 5. Here, I review the theory and the evidence on the three most relevant explanations that have been proposed for the agglomeration of economic activity: (1) advantages deriving from thick labor markets; (2) advantages deriving from proximity to providers of intermediate non-tradable goods and services; (3) localized knowledge spillovers.[41]

4.2.1. Thick labor markets

Thick markets have long been understood to be more attractive than thin markets when frictions of some type separate demand from supply. For example, in the influential barter model by Diamond (1982), the probability of finding a trading partner depends on the number of potential partners available, so that an increase in the thickness of the market makes trade easier. This generates multiple steady state equilibria and, in each of them, the equilibrium level of production is not efficient. In the context of local labor

---

[40] They use a generalization of the Melitz and Ottaviano (2008) model to distinguish between agglomeration economies and firm selection. Using French data on manufacturing establishments, they find that firms in large cities are more productive than firms in small cities, but find no difference in the amount of selection between small and large cities.

[41] Other explanations have been proposed for agglomeration economies, but they seem to be less relevant than the three described above. Some concentration of economic activity may be explained by the presence of natural advantages that constrain specific productions to specific locations. In practice, while natural advantages may be important in some industries, they are unlikely to be important determinants of agglomeration in most industries. In an often cited paper, Ellison and Glaeser (1999) show that natural advantages matter in some cases, but they account for only 20% of the observed degree of agglomeration. Historically, proximity of firms to consumers might also have played a role in the agglomeration of economic activity in metropolitan areas (Krugman, 1991b). In practice, however, the substantial decline in transportation costs makes this explanation less relevant for most industries at the present time. Glaeser and Kohlhase (2003) calculate that the cost of moving goods by rail or trucking has declined more than 90 percent over the last century.

markets, there are two reasons why workers and firms may find thick labor markets in large metropolitan areas attractive: better matches and lower risk.[42]

First, in the presence of worker and firm heterogeneity, a worker–firm match may be more productive in areas where there are many firms offering jobs and many workers looking for jobs. The higher quality of the worker–firm match in thicker labor markets may result in higher productivity and higher wages. This notion was first formalized twenty years ago in a model by Helsley and Strange (1990). Acemoglu (1997) and Rotemberg and Saloner (2000) propose variants of this hypothesis. In Acemoglu's model, employers in thick labor markets invest in new technologies because they know they can find specialized employees. At the same time, employees in thick markets invest in human capital because they know that when they change jobs, their human capital will be valued. Rotemberg and Saloner capture a similar idea in a model with multiple cities with identical factor endowments. In their setting, agglomeration of production is caused by the fact that having more competition between firms to hire skilled workers makes it easier for skilled workers to recoup the cost of acquiring industry-specific human capital.

Note that the relevant definition of labor market thickness is likely to depend not just on the size of a metropolitan area, but also on the skill set of a given worker. Two workers with different skills living in the same metropolitan areas may be exposed to vastly different labor market thickness. For example, a bioengineer and an architect living in the same city may face different market thickness, depending on the local agglomeration of bioengineering firms and architectural firms.

A second potential benefit of thick labor markets is the provision of insurance to workers and firms against idiosyncratic shocks. Thick labor markets reduce the probability that a worker remains unemployed following an idiosyncratic negative product demand shock to her employer. The presence of a large number of other employers implies a lower probability of not finding another job. At the same time, thick labor markets reduce the probability that a firm can't fill a vacancy, following an idiosyncratic shock to the labor supply of an employee. The presence of a large number of workers ensures a lower probability of not finding another worker. As in the case of match quality, this argument applies particularly to workers with specialized skills.

These two versions of the thick labor market hypothesis have different empirical implications. If the size of the labor market leads to better worker-firm matches, we should see that firms located in denser areas are more productive than otherwise identical firms located in less dense areas. The fact that the size of the labor market leads to a lower risk of unemployment for workers and a lower risk of unfilled vacancies for firms, does not imply differences in productivity between dense and less dense areas, but differences in wages. The sign of these wage differences is unclear a priori, because it depends on the magnitude of the compensating differential that workers are willing to pay for lower risk

---

[42] Labor market pooling externalities were first proposed by Marshall (1890).

of unemployment (generated by an increase in labor supply in denser areas) relative to the cost savings that firms experience due to lower risk of unfilled vacancies (generated by an increase in labor demand in denser areas). The former can be thought of as an increase in labor demand in thicker markets, while the latter can be thought of as an increase in labor supply in thicker markets.

Although the idea that thick labor market are beneficial is an old one, the existing empirical evidence is still limited. This is due in part to the fact that the quality of a worker–firm match has proven difficult to measure in practice.[43] Using a semi-structural model of the labor market, Petrongolo and Pissarides (2005) provide one of the earliest tests of the scale effect in job search based on the comparison of the number of job matches in labor markets of different sizes. The idea of the test is very simple. If the total number of vacancies and unemployed workers in a city is twice as large as the total number of vacancies and unemployed workers in another city, we should see that the number of matches per unit of time in the former city is more than twice as large the number of matches in the latter. Using British data, Petrongolo and Pissarides compare the offer arrival rate and the wage offer distribution in London (a thick labor market) with the rest of the country. They find significant scale effects in wage offers, but not in actual matches. However, the lack of observed scale effects in matches could in part be explained if the reservation wage endogenously adjusts to the size of the market.

A second piece of empirical evidence has to do with the relationship between frequency of job changes and size of the labor market. Wheeler (2008) and Bleakley and Lin (2006) find evidence that, although indirect, is generally consistent with localized increasing returns to scale in matching. In particular, using longitudinal data from the NLSY, Wheeler (2008) documents that the probability of changing industry is positively correlated with the size of the labor market for young workers and negatively associated with the size of the labor market for older workers. Bleakley and Lin (2006) find similar results for industry and occupation changes using cross-sectional data from the Census. They also find that this pattern remains true even in the case of involuntary separations. In other words, early in a career, when presumably workers are shopping around for a good match, industry and occupation changes occur more often in large, diverse local markets than in small, specialized ones. Later in a career, when changing industry or occupation becomes more costly because it may involve giving up specialized skills, industry and occupation changes occur less often in large markets, presumably because matches are better. The existing evidence, while generally consistent with the notion that larger labor markets facilitate matching, is still indirect. A more direct test might involve the duration of the match, as a measure of the quality of the match (Jovanovic and Rob, 1989). A testable implication is that job duration should be longer in larger labor markets than in smaller markets.

[43] In this respect, Puga (2009) argues that "the increasing availability of matched employer employee microdata will encourage more work on agglomeration through matching".

An additional testable implication of this hypothesis is that the productivity benefit of thick labor markets should be particularly important for industries that rely on specialized labor. Consider, for example, a digital media software engineer. If digital media firms have heterogeneous technologies, and digital media software engineers have heterogenous skills, it is likely that the match between a worker's specific skills and a firm's specific technology is more productive in a city where there are many digital media firms than in a city where there is only one digital media firm.[44] By contrast, the thickness of the labor market may not significantly improve the match quality and the productivity of less specialized workers in the same city, like manual laborers. An empirical test could involve estimating the difference in the correlation between measures of match quality (for example, wages and job duration) and market size, for workers who live in the same city but have different skill levels (specialized vs non specialized) and work in different industries (locally agglomerated vs locally non–agglomerated).

Fallick et al. (2006) posit that in high-tech industry clusters like Silicon Valley, high job mobility may facilitate the reallocation of resources towards firms with superior innovations, but it may also create human capital externalities that reduce incentives to invest in new knowledge. They argue that in the computer industry, the innovation benefits of job-hopping exceed the costs from reduced incentives to invest in human capital, while in other parts of the high-tech sector the opposite might be true. Their evidence is consistent with this notion, but it is too indirect to be definitive.[45]

The evidence in Greenstone et al. (forthcoming) also is consistent with the notion that spillovers occur between firms that share a similar worker pool. The magnitude of the productivity spillovers that they uncover depends on the economic linkages between the new plant and the incumbent plant. In their data, spillovers are larger for pairs of firms that belong to industries that share the same set of workers. This lends credibility to the notion that labor market pooling is an important source of agglomeration economies.

Costa and Kahn (2000) point out that thick labor markets are particularly important for dual career households, because thick labor markets may solve the co–location problem. The economic return of being in a large labor market relative to a small market is increasing over time and Costa and Kahn attribute at least half of the increased agglomeration of skilled workers in large cities to the growing severity of the co–location problem.

In his influential book on economic geography, Krugman (1991a,b) proposes an alternative version of the thick labor market hypothesis. He argues that an advantage

---

[44] Indirect support for this hypothesis can be found in Baumgardner (1988), who finds that doctors perform a narrower and more specialized set of activities in large metropolitan areas than in small metropolitan areas.

[45] Andersson et al. (2007) show that thicker urban labor markets are associated with more assortative matching between workers and firms and argue that production complementarity and assortative matching are important sources of the urban productivity premium. Using a matched employer–employee database for Italy, Mion and Naticchioni (2009) also find important amount of assortative matching.

of thick labor markets is that idiosyncratic demand shocks to firms are less likely to affect equilibrium wages. To see why it may matter, consider a firm experiencing an idiosyncratic positive productivity shock. If the labor market is thick, this firm faces a relatively flat supply of labor. If the labor market is thin, the firm faces an upward sloping supply of labor and the firm-specific shock may ultimately result in higher labor costs. Consistent with this hypothesis, Overman and Puga (forthcoming) show that industries characterized by more idiosyncratic volatility are more spatially concentrated.

In sum, thickness of the labor market is a potentially promising explanation for the agglomeration of economic activity in metropolitan areas. It is highly plausible that workers prefer to be in areas with thick labor markets because of an increased probability of a better match with an employer and a reduction in the probability of unemployment. At the same time, it is also highly plausible that firms prefer to be in areas with thick labor markets because of an increased probability of a better match with an employee and a reduction in the probability of unfilled vacancies. While intriguing, most of the existing empirical evidence is still quite indirect. This clearly is an area that should receive increased attention by labor economists in the years to come.

4.2.2. Thick market for intermediate inputs

A second possible explanation for agglomeration economies centers on the availability of specialized intermediate inputs. Concentration of specialized industrial production can support the production of non-tradable specialized inputs. The agglomeration economy in this case is generated by the sharing of inputs whose production is characterized by internal increasing returns to scale. This explanation is likely to be particularly relevant for firms that utilize intermediate inputs that are both highly specialized and non-tradable. Consider, for example, an industry where production crucially depends on the availability of specialized local producer services, such as specialized repair services, engineering services, venture capital financing, specialized legal support, or specialized shipping services. To the extent that these services are non-tradable—or are costly to deliver to distant clients—new entrants in this industry have an incentive to locate near other incumbents. By clustering near other similar firms, entrants can take advantage of an existing network of intermediate inputs suppliers. In equilibrium, cheaper, faster or more specialized supply of intermediate goods and services makes industrial clusters attractive to firms, further increasing the agglomeration. This concentration process will go on up to the point where the increase in land costs offsets the benefits of agglomeration.

While this idea has been around for a long time, the first to formalize it are Abdel Rahman and Fujita (1990), who propose a model where final goods are tradable, but intermediate inputs are non-tradable and are produced by a monopolistically competitive industry. In the model, firms that locate in dense areas share a larger and wider pool of intermediate inputs suppliers, while otherwise similar firms that locate in rural areas share a smaller and narrower pool of intermediate inputs suppliers. This difference generates

agglomeration advantages because an increase in the number of firms in an area results in a wider local supply of inputs and therefore in an increase in productivity.

The evidence in Holmes (1999) offers direct support for the input sharing hypothesis. Using data on manufacturing plants, he documents that manufacturing establishments located in areas with many other establishments in the same industry make more intensive use of purchased intermediate inputs than otherwise similar manufacturing establishments in areas with fewer establishments in the same industry. Notably, this relationship only holds among industries that are geographically concentrated. Spatial proximity has limited impact on geographically dispersed industries.

Building on an idea first proposed by Rosenthal and Strange (2001), Overman and Puga (forthcoming) provide an alternative test of this hypothesis by relating measures of geographic concentration for each industry to industry-specific measures of the importance of input sharing. They find support for the notion that the availability of locally supplied inputs is an important empirical determinant of industrial clusters.

Ellison et al. (forthcoming) propose an alternative approach to the one taken by Rosenthal and Strange (2001) and Overman and Puga (forthcoming). They seek to understand the mechanics of the agglomeration process by focusing on how industries are coagglomerated. Different agglomeration theories have different predictions about which pairs of industries should coagglomerate. For example, if input markets are important, then firms in an industry should be observed to locate near industries that are their suppliers. On the other hand, if labor market pooling is important, then industries should locate near other industries that employ the same type of labor. Ellison, Glaeser and Kerr find evidence that input–output dependencies, labor pooling and knowledge spillovers are all significant determinants of agglomeration, but input–output dependencies appear to be empirically the most important channel.

4.2.3. Knowledge spillovers

Economists have long speculated that the agglomeration of human capital may generate positive spillovers, over and above its private effect.[46] Different explanations have been offered for such spillovers. For example, physical proximity with educated workers may lead to better sharing of ideas, faster innovation or faster technology adoption. Perhaps the most influential theoretical contribution in this area is the model by Lucas (1988). In that paper, human capital is assumed to have two effects. First, an individual's own human capital has the standard effect of increasing her own productivity. Second, the average aggregate level of human capital contributes to the productivity of all factors of production. This second effect is an externality, because "though all benefit from it, no individual human capital accumulation decision can have an appreciable effect on average human capital, so no one will take it into account" in deciding how much to invest in

---

[46] This question has important implications for education policies. The magnitude of the social return to education is important for assessing the efficiency of public investment in education.

human capital accumulation. In Lucas' view, human capital externalities may be large enough to explain long–run income differences between rich and poor countries. While in the model the externality is simply built into the production function in black-box fashion, Lucas posits that the sharing of knowledge and skills through formal and informal interaction may be the mechanism that generates the externality. More recent models build on this idea by assuming that individuals augment their human capital through pairwise meetings with more skilled neighbors at which they exchange ideas. Examples include Glaeser (1999),[47] Jovanovic and Rob (1989), Henderson and Black (1999), Jaffe et al. (1993), Duranton and Puga (2001) and Saxenian (1994).[48]

A second class of models explains positive human capital externalities as pecuniary externalities that arise because of search or endogenous skill–biased technical change. Consider for example the case proposed by Acemoglu (1996), where job search is costly, and physical and human capital are complements. The privately optimal amount of human capital depends on the amount of physical capital a worker expects to use. The privately optimal amount of physical capital depends on the supply of human capital. If a group of workers in a city increases its level of education, firms in that city, expecting to employ these workers, would invest more in physical capital. Since search is costly, some of the workers who have not increased their education would end up working with more physical capital and hence earn more than similar workers in other cities. As in Lucas, the presence of skilled workers in a city generates external benefits for other workers there. But what distinguishes Acemoglu's story from Lucas' story is that this result does not follow from assumptions on the production function, but rather is derived from market interactions. Even though all the production functions of the economy exhibit constant returns to scale in Acemoglu, the complementarity of human capital and physical capital, coupled with frictions in the job search process, generate a positive relationship between the average wage and average human capital, holding constant workers' individual human capital.[49]

---

[47] Glaeser (1999) argues that young workers move to cities because interactions with experienced workers help them increase their human capital.

[48] Of course, it is also possible that human capital externalities are negative. If education functions as a signal of productive ability, rather than enhancing productivity directly, the private return may exceed the social return. This is a case where people with higher innate ability signal their higher innate productivity by enduring extra years of schooling. If schooling is more difficult for individuals with low innate productivity than for individuals with high innate productivity, then, even if schooling itself is worthless in terms of enhancing productivity, it still may be a useful screening device for employers to identify more productive job applicants. In this case, increasing the average schooling in a city would result in an increase in aggregate earnings that is smaller than the private return to schooling.

[49] Although differences across cities in their *quantity* of physical capital play a central role in this model, differences in the *quality* of physical capital (technology) could arguably generate similar conclusions. Specifically, if skills and technology are complementary, it is plausible to assume that the privately optimal amount of human capital depends not only on the amount of physical capital a worker expects to use, but also on the technological level that characterizes such capital. Similarly, in models with endogenous skill-biased technical change, an increase in the supply of educated workers increases the size of the market for skill-complementary technologies and stimulates the R&D sector to spend more effort upgrading the productivity of skilled workers (Acemoglu, 1997).

There is growing empirical evidence that human capital spillovers and knowledge spillovers may be particularly important in certain high-tech industries, where innovation has been shown to be linked to human capital externalities. Because human capital spillovers and knowledge spillovers are invisible, most empirical studies resort to indirect evidence to test for the presence of spillovers. Jaffe et al. (1993) are an exception, in that they provide direct evidence of spillovers using a "paper trail" based on patent citations to test the extent to which knowledge spillovers are geographically localized. Because patents are publicly available, in the absence of localized spillovers, citations should not depend on the location of the inventor. The key empirical challenge of the paper is to distinguish between geographic patterns of patent citations caused by spillovers from patterns caused by other sources of agglomeration effects. To address this issue, the authors construct "control" samples of patents that have the same temporal and technological distribution as the patent citations and compare the two patterns of geographic concentration. They find that references to existing patents that inventors include in their patent applications are likely to come from the same state or metropolitan area as the originating patent application. They conclude that patent citations are geographically localized and that knowledge spillovers appear to be large.

In a related study, Zucker et al. (1998) argue that the presence of specialized human capital is the main determinant of the localization of biotechnology firms in the US. They show that the stock of human capital of scientists in certain cities—measured in terms of the number of publications reporting genetic-sequence discoveries in academic journals—is correlated with the location of new biotech firms. This effect may reflect, at least in part, human capital externalities, because it is not just a reflection of the presence of universities and government research centers in areas where outstanding scientists are located. The introduction of new products—spatially clusters more in industries where new knowledge plays a more important role, holding constant the degree of spatial clustering of economic activity.[50] In recent work, Carlino et al. (2007) use patents to measure innovation, and find that the number of patents per capita is positively correlated to the density of employment in the highly urbanized portion of metropolitan areas. A city with twice the employment density of another city has 20% more patents per capita. Local human capital appears to be the most important predictor of per capita rates of

---

[50] On the other hand, the exact magnitude of the spillovers is still debated. Audretsch and Feldman (1996) use data on the IPO's of biotech firms to link the location of the biotechnology firm with the location of the university-based scientists affiliated with the firm. They conclude that "while proximity matters in establishing formal ties between university-based scientists and companies, its influence is anything but overwhelming". In earlier work, Adams and Jaffe (1996) study the composition of the knowledge transfers within and across firms. They use manufacturing plant-level data to examine the productivity effects of R&D performed in a plant, outside a plant but inside the parent firm that owns the plant, and in external plants in the same geographic area or industry. They find that R&D of other firms in the same industry is correlated with a plant's own productivity, holding industry constant. However, identification is based on questionable assumptions and the potential for omitted variable bias makes it hard to draw firm conclusions about causality.

patenting. However, due to the absence of a credible research design, this study fails to establish causality.

These earlier studies are consistent with the notion that knowledge spillovers may be important in a limited number of high-tech or high-skill industries. More recent studies seek to provide more general—and often better identified—tests for human capital spillovers.

Using estimates of establishment-level production functions, Moretti (2004a) shows that manufacturing plants are significantly more productive in cities with higher human capital, holding constant individual plant human capital. The magnitude of spillovers between plants in the same city declines in economic distance as in Greenstone et al. (forthcoming). Much of the estimated spillover comes from high-tech plants. For non high-tech producers, the spillover appears to be limited. Consistent with the predictions of the spatial equilibrium model, the productivity gains uncovered by Moretti appear to be offset by increased labor costs. The estimated productivity differences between cities with high human capital and low human capital are similar to observed differences in wages of manufacturing workers, indicating an almost complete offset. While the documented productivity gains from human capital spillovers are statistically and economically meaningful, the implied contribution of human capital spillovers to economic growth does not appear to be large. Moretti estimates that human capital spillovers were responsible for an average of a 0.1% increase in output per year during the 1980s.[51]

A recent paper by Beaudry et al. (2006) proposes a hypothesis that may potentially explain Moretti's findings. They argue that over the past 30 years, technological change resulted in increases in the productivity of skilled workers in cities that had many educated workers. The estimates support the idea that differences in technology use across cities (measured by the adoption of computers) and its effects on wages reflect an equilibrium response to local factor supply conditions. In particular, cities initially endowed with relatively abundant and cheap skilled labor adopted computers more aggressively than cities with relatively expensive skilled labor, causing returns to skill to increase most in cities that adopted computers most intensively.

A growing number of studies focus on the effect of aggregate human capital on earnings. A simple framework indicates that increases in the aggregate level of human capital in a city have two distinct effects on wages. First, imperfect substitution between educated and uneducated workers indicates that an increase in the number of educated workers will lower the wage of the educated and raise the wage of uneducated workers. Second, human capital spillovers may raise the wage of both groups. Imperfect substitution and spillovers both increase wages of uneducated workers, while the impact of an increase in the supply of educated workers on their own wage is determined

---

[51] Glaeser et al. (1995) report that income per capita has grown faster in cities with high initial human capital in the post-war period.

by two competing forces: the first is the conventional supply effect which makes the economy move along a downward sloping demand curve; the second is the spillover that raises productivity. Using metropolitan areas as a definition of local labor markets, Rauch (1993) and Moretti (2004b), among others, document that wages are significantly higher in metropolitan areas with higher human capital, holding constant individual worker human capital. In particular, consistent with a model that includes both conventional demand and supply factors as well as human capital spillovers, Moretti finds that a one percentage point increase in the labor force share of college graduates increases the wages of high-school drop-outs, high-school graduates and college graduates by 1.9%, 1.6% and 0.4%, respectively. Using states as a definition of local labor markets, Acemoglu and Angrist (2000) fail to find significant evidence of human capital spillovers. A possible explanation of this difference is the evidence in recent work by Rosenthal and Strange (2008). They find that proximity to college graduates is associated with increases in wages but that these effects attenuate sharply with distance. If this is the case, it is possible that states are too large of geographic units to allow for the detection of human capital spillovers.[52]

The findings in Glaeser and Mare (2001) described above are consistent with a model where individuals acquire skills by interacting with one another, and dense urban areas increase the probability of interaction. In a related paper, Peri (2002) shows that young educated workers receive a lower wage premium in urban areas than their old educated workers, but in spite of this, young educated workers are overrepresented in urban areas. Peri argues that learning externalities are an important explanation. Workers learn from each other when they are young, so living in dense urban areas may raise human capital accumulation more than living in a rural area. The negative compensating differential indicates that young workers value such human capital externalities. As they grow older, the importance of knowledge spillovers diminishes, and some of them move toward non-urban areas.

Some studies have posited that areas with a more educated populace are more likely to generate new business ideas and new firms. This is not a market failure. However, if skilled people are more likely to innovate in ways that employ other skilled people, this creates an agglomeration economy where skilled people want to be around each other. Berry and Glaeser (2005) present evidence consistent with a model of urban agglomeration where the number of entrepreneurs is a function of the number of skilled people working in an area. Consistent with this hypothesis, Doms et al. (2009) use a new panel of startup firms to show that areas that possess more skilled labor also possess higher rates of self-employment and more skilled entrepreneurs. Moreover, conditional

---

[52] Additionally, Acemoglu and Angrist estimate spillovers coming from high school graduation. On the contrary, Moretti and Rosenthal and Strange identify spillovers coming from variation in college graduation. It is possible that an increase in the number of those who finish high-school has a different external effect than an increase in the number of those who go to college.

on owner's education, higher education levels in the local market are positively correlated with improved business outcomes.

5. IMPLICATIONS FOR POLICY

The empirical evidence surveyed in Section 4 points to the concrete possibility that agglomeration of economic activity generates significant economies of scale at the local level. It is therefore natural to raise the question of the desirability of government intervention. In a world with vast disparities in income levels across localities and with significant agglomeration externalities, what is the proper role of economic policy? Should national or local governments provide subsidies to firms to locate in their jurisdiction?

   In this section, I discuss the economic rationales for location-based policies. I define location-based policies as government interventions aimed at reallocating resources from one location to another location. These policies are widespread both in the US and in the rest of the world. In the US, state and local governments spend $30-40 billion per year on these policies, while the federal government spends $8-12 billion (Bartik, 2002). Examples of location-based policies typically adopted by local and state governments include direct subsidies and/or tax incentives for local firms, subsidized loans, industrial parks, technology transfer programs, export assistance and export financing, the provision of infrastructure, workforce training, subsidies to higher education and area marketing. Bartik (1991) provides a comprehensive taxonomy and discussion of the different types of policies. More generally, states and cities compete based on income and corporate tax rates, labor and environmental regulations, and many other forms of intervention that affect the relative profitability for firms of locating in each jurisdiction. The US federal government also promotes several location-based policies. The Tennessee Valley Authority and the Appalachian Regional Commission are important historical examples of large federal programs that target poor rural areas for development aid. A more recent example is the Federal Empowerment Zones Program, which is a system of subsidies for businesses located in poor urban neighborhoods. Location-based policies are also widespread in Europe in the form of European Union regional transfers; and in Asia, in the form of special economic zones.[53]

   In general, economists think that there are two possible rationales for government intervention in the economy: equity and efficiency. Location-based policies are no exception. I begin in Section 5.1 by discussing several aspects of the equity rationale. In Section 5.2, I turn to the efficiency rationale. Bartik (1991, 2002) and

[53] A prominent example of the latter is the successful program adopted by the government of Taiwan to subsidize R&D in semi-conductors and other high-tech fields.

Glaeser and Gottlieb (2009) provide authoritative discussions of the economic rationales for local subsidies with somewhat differing conclusions on their desirability.[54]

5.1. Equity considerations

5.1.1. Incidence of subsidies

It is tempting for policymakers to support policies intended to help disadvantaged areas. The main argument in favor of these policies is that by helping disadvantaged areas, the government helps disadvantaged individuals. The spatial equilibrium model outlined in Section 3 suggests that this argument is at least in part flawed, since the poor are unlikely to fully capture the benefits of location-based subsidies. In a world where workers are mobile, targeting locations instead of individuals is an ineffective way of helping disadvantaged individuals.

To see this more concretely, I consider first the case where subsidies for firms to locate in a given locality are financed by the central government, and then the case where subsidies are financed locally.

(a) Centrally financed subsidies Consider a location-based redistributive policy intended to help disadvantaged areas. Assume for example that the central government taxes residents of areas with high (nominal) income to provide subsidies to firms to locate in areas with low (nominal) income. Most countries have these types of redistributive policies. An example of this policy in the US is the Empowerment Zones Program mentioned above. Another example is represented by the ubiquitous state policies designed to attract businesses to poor parts of their jurisdiction. In Canada, an equalization program transfers income from high income provinces to low income provinces. The European Union has a similar transfer program aimed at transferring EU development funds to regions with below average income.[55]

The incidence of this policy and its redistributive implications ultimately depend on the elasticity of local labor supply and housing supply. Section 3 clarifies that if workers are highly mobile, they will arbitrage some or all of the benefits associated with this transfer by relocating to the area favored by the transfer, thus bidding up the price of housing. In the case of high elasticity of local labor supply and less than infinite elasticity of housing supply, this increase in housing prices will offset most of the welfare gains that might otherwise accrue to existing residents.[56] In the extreme case of perfect mobility,

---

[54] Boarnet and Bogart (1996), Goolsbee and Maydew (2000), Faulk (2002), Bennmarker et al. (2009), Bondonio and Engberg (2000), Peters and Fisher (2002), Greenstone and Moretti (2004), Greenstone et al. (forthcoming) provide recent estimates of the employment effects of local subsidies. The last two papers have a discussion of the policy implications of these subsidies. Work by Bartik (1991) and Papke (1993, 1994) represent important early contributions that helped to frame the policy debate on location-based policies in economic terms. See also Wasylenko (1997).

[55] Similarly, one can think of a direct subsidy to residents of poor areas. For example, the national government could tax residents of areas with high (nominal) income to provide a transfer to residents of areas with low (nominal) income. This type of redistribution has similar implications.

[56] Of course, an infinitely elastic supply of housing would prevent this price increase.

increases in the price of housing fully offset the transfer. In this setting, location-based redistributive policies intended to help areas with low nominal income have virtually no effect on the utility of workers. The only beneficiaries of this policy are landowners in the targeted areas. Effectively, this policy amounts to a transfer of wealth from landowners in the rest of the country to landowners in the targeted areas. If the theoretical model is a fair approximation of the real world, then its basic premises cast considerable doubt on the desirability of redistributing income from areas with high income to areas with low income.[57]

How does this conclusion change if labor is not highly mobile? The model in Section 3 indicates that if individuals have significant preferences over specific locations, labor is less mobile. In this case, the *marginal* worker is indifferent across locations, but the *average* worker is not and location-based redistributive policies have the potential to affect the utility of the average worker. In particular, inframarginal workers in rich areas experience an increase in taxes and a decrease in the cost of living, while inframarginal workers in poor areas experience an increase in transfers and an increase in the cost of living. Overall, the redistributive effect is complicated and unlikely to be clear ex-ante, because it crucially depends on individual preferences for location, which are unlikely to be observed by policy makers. This lack of observability makes it difficult to implement policies of this type in practice.

Busso et al. (2009) provide the first comprehensive empirical welfare analysis of a location-based policy, namely the Empowerment Zones program. Using a remarkably detailed series of data from the Census of Population, the Longitudinal Business database and the Standard Statistical Establishment List, together with an identification strategy based on areas that applied for the credits but did not receive it, they are able to credibly quantify the incidence and deadweight loss of the program. Consistent with the spatial equilibrium model outlined in Section 3, they find that both wages and housing values increase significantly in the neighborhoods that benefit from the federal subsidy relative to the counterfactual neighborhoods in rejected zones. This increase is consistent with an upward sloping local supply of labor.

In terms of incidence, Busso et al. (2009) find that the program unambiguously benefits landowners in Empowerment Zone areas. This increase in housing values is not particularly surprising. Together with the failure to find any changes in overall area population, the increase in housing values indicates that the supply of housing is inelastic, at least in the short run. More surprisingly, the program also benefits workers who reside in the area, since they experience an increase in nominal wages larger than the increase in housing costs. Based on the model in Section 3, this finding is consistent with the

---

[57] In some cases, these types of policies may even have perverse consequences for targeted localities. Since the early 1970's, the Canadian Unemployment Insurance program—a federal program—has been regionally differentiated, with more generous benefits in high unemployment areas. There is considerable evidence that this feature has had significant undesirable side effects. See Kuhn and Riddell (2010), for example.

presence of significant locational preferences on the part of residents. Workers who live in the targeted areas appear to have strong preferences for their current residence, so that their mobility is limited. This is notable, because the treated areas are neighborhoods within much larger metropolitan areas, and it is in principle possible to commute into the Empowerment Zone areas from the rest of the metropolitan area without having to change residence. Given the finding of small deadweight losses, the overall welfare assessment of the program appears encouraging.

Busso et al. (2009) also find that the provision of Empowerment Zone subsidies results in an increase in the productivity of local firms. While it is difficult to identify the exact channel for this productivity increase, three plausible candidates are: an improvement in public infrastructure or other local public goods; some form of agglomeration economy; the role of the subsidies as a coordination mechanism for private investment. Because of the nature of the subsidy and the type of production that is common in empowerment zone areas, the most plausible candidate appears to be the last one.

(b) Locally financed subsidies I consider now the case where local taxpayers bear the cost of subsidizing firms that locate in their jurisdiction. The welfare effects of this type of policy depend on how similar the locations competing to attract new firms are. In particular, in the case of homogenous locations, a locally financed subsidy has no effect on residents' welfare since all the rent associated with the subsidy is transferred to the firm. In the case of heterogenous locations, a locally financed subsidy will benefit landowners by an amount proportional to the difference in production cost between the location with the most desirable attributes and the location with the second most desirable attributes. The effect on workers depends on the importance of their preferences for location. Limited preferences for locations imply high mobility and therefore limited welfare changes. Significant preferences for locations imply low mobility and therefore significant welfare changes for inframarginal individuals.

To see this, I use the framework proposed by Greenstone and Moretti (2004). I assume that local governments bid to attract new firms to their jurisdiction by offering subsidies, and that the cost of the subsidy is financed by increases in local property taxes, and therefore is capitalized into land values. Let V_{ij} denote the benefit of the increase in the level of economic activity generated by new firm j for locality i, assumed to be known to all the other localities.[58] Unlike in the case examined in the previous subsection—where the subsidy was paid for by the central government—here a successful bid now involves a trade-off for a locality. Let C_{ij} denote the direct monetary cost of the subsidy. The partial equilibrium change in welfare for the winning county can be expressed as $V_{ij} - C_{ij}$.

---

[58] This is equal to the change in utility for local residents and landowners. If labor is mobile, workers are always indifferent, and landowners are the only set of agents whose welfare may be affected. If workers have preferences for location, however, both workers and landowners are affected. While this type of subsidy is quite common in practice, it is not always obvious what localities seek to maximize in this process. One possibility is that, in the presence of unemployment, localities seek to maximize job creation. My theoretical model in Section 3 is a full employment model that by construction rules out unemployment.

Let the value to firm j of locating in county i be Z_{ij}. Due to differences in technology, the same locality may be more or less attractive to different firms.[59] A higher Z_{ij} implies that the production costs of firm j are lower in locality i. I assume that Z is known to all localities. The total value for a firm of locating in a particular locality is the sum of the subsidy and the county-specific cost advantages. A firm will select the locality where this sum, $B_{ij} + Z_{ij}$, is maximized. In order to obtain the highest subsidy, I assume that the firms conduct an English auction in the presence of independent, or private, values. I further assume that there is no collusion in the bidding among counties.

Consider the case where counties are homogeneous in V and Z: $V_{ij} = V_0$ and $Z_{ij} = Z_0$ for all i. In this case, the firm simply chooses the location that offers the highest subsidy, B. The equilibrium bid, B^*, is $B^* = V_0$. This implies that successfully attracting the firm does not change residents' welfare: $V_0 - B^* = 0$. The reason is that each jurisdiction keeps raising its bids until it is indifferent between winning and losing, so that the equilibrium bid is such that the entire economic rent is transferred to the firm. In this case landowners are indifferent, since the benefit of the new firm is fully offset by the increase in property taxes. This result is similar to the result in the tax competition literature where local jurisdictions keep taxes on capital low because of a fear of capital flight.

Consider now the more general case where counties' valuations of attracting the plant and plants' valuations of counties are heterogenous. Assume for simplicity that there are only two locations, high V (V_H) and low V (V_L); and two levels of Z, high Z (Z_H) and low Z, (Z_L). If V and Z are positively correlated, the location with high V also has high Z. In this case, this location gains the most from attracting firm j and it is also the most attractive to firm j. The optimal bid is such that the firm is indifferent between moving to either city. Unlike the homogenous case, where all the economic rent is bid away, in this case, the H county enjoys an economic rent that comes from the fact that it has characteristics that are desirable to the firm.

If labor is highly mobile, and housing is inelastically supplied, this economic rent will be capitalized into land values. In particular, land values increase by an amount $(V_H - V_L) + (Z_H - Z_L) > 0$, proportional to the difference in V and the difference in Z. The same intuition applies to the case where V and Z are negatively correlated, and location 1 has high V and low Z, while location 2 has high Z and low V. If $V_H + Z_L > V_L + Z_H$, location 1 wins the firm by bidding an amount B^* that makes the firm indifferent: $B^* = V_L + (Z_H - Z_L)$. As in the case of positive correlation between Z and V, the winning location enjoys a rent that is capitalized in land values, although the rent is lower than the rent in the case of positive correlation: $(V_H - V_L) - (Z_H - Z_L) > 0$. A similar conclusion applies if $V_H + Z_L < V_L + Z_H$. In this case location 2 is the winner and its land prices increase by $(Z_H - Z_L) - (V_H - V_L) > 0$.

---

[59] For example, the presence of a harbor, an airport or a freeway may be more important for some productions than others. Similarly, the presence of stringent environmental or labor regulations may affect some firms more than others.

5.1.2. Taxes and transfers based on nominal income

An important redistributive implication of the spatial equilibrium model has to do with federal taxation and federal transfers. Federal taxes and transfers are calculated based on nominal income. By setting taxes and federal transfers in nominal terms, the federal government engages in a hidden form of location-based redistribution, because workers with the same real income pay higher federal taxes in high-cost areas than in low-cost areas. Albouy (2009) estimates that workers in cities where nominal wages are above the national average pay up to 27% more in federal taxes than similar workers in cities where nominal wages are below the national average. As a consequence, $270 billion each year are transferred from areas with high nominal wages to areas with low nominal wages.

In equilibrium, if workers are mobile, wages and land prices should adjust to compensate workers. However, the resulting geographic distribution of employment is inefficient, since it penalizes highly productive cities and favors less productive cities. In other words, this policy artificially lowers economic activity and property values in cities where labor is more productive and nominal wages are higher. At the same time it increases economic activity and property values in cities where labor is less productive and nominal wages are lower. The net result is a loss in overall welfare. Albouy calculates that the long-run employment loss in high nominal wage areas is about 13%, while the loss in land and housing values is about 21% and 5%, respectively. Albouy suggests that one solution is to make taxes independent of where workers live so that they are effectively lump sum location-wise.

A related problem arises when thinking about transfer payments. Should they be based on nominal or real income? Using a spatial equilibrium model similar to the one in Section 3, Glaeser (1998) derives the conditions under which welfare payments should be adjusted for differences in the local cost of living. He concludes that the optimal transfer depends on mobility and preferences for amenities. In the case of perfect mobility, transfer payments that correct for differences in the local cost of living are inefficient, because they end up being capitalized in the price of land, further raising land costs in expensive areas. With limited mobility, a correction for local cost of living differences is optimal under the assumption that amenities and income are complements.

5.1.3. Nominal and real differences across skill groups and regions

The spatial equilibrium model has implications for how one should measure earnings and income differences between skill groups or between regions. In most countries, there are large cost of living differences across regions. These differences are typically largely driven by differences in the price of land. When comparing earnings or income across skill groups or across regions, the question arises of whether nominal measures should be used or whether real measures should be used. This question matters because the magnitude of income differences between skill groups or between regions has implications for the desirability of redistributive policies.

Earnings differences between skill groups Consider the increase in earnings inequality in the US labor market in the past three decades. As documented by a large literature in labor economics, starting in 1980 the nominal earnings of skilled workers have grown significantly faster than the nominal earnings of unskilled workers. In the same period, there have been increasing differences in the geographic distribution of skilled and unskilled workers. Skilled workers have increasingly concentrated in cities with high costs of land, while unskilled workers have increasingly concentrated in cities with low costs of land. This geographic sorting suggests that skilled workers might have experienced higher increases in the cost of living.

Consistent with this hypothesis, Moretti (forthcoming) finds that earnings inequality measured in *real* terms has grown significantly less than earnings inequality in *nominal* terms. The model in Section 3 shows that the implications for utility inequality depend on the underlying reasons for the geographic sorting. If the sorting of skilled workers into more expensive cities is driven by increases in the relative demand of skilled labor in those cities, the increase in relative utility of skilled workers is smaller than the increase in their nominal wage. On the other hand, if the sorting of college graduates into more expensive cities is driven by an increase in the relative supply for skilled labor in those cities (i.e. an increase in the attraction of local amenities), increases in the cost of living in these cities reflect the increased attractiveness of the cities for skilled workers. In this case, there may still be a significant increase in utility inequality even if the increase in real wage inequality is limited. Moretti (forthcoming) argues that the evidence is more consistent with the notion that shifts in the relative demand for skilled labor are the main force driving changes in the number of skilled workers across metropolitan areas and that the increase in well-being inequality is smaller than the increase in nominal wage inequality. These results are related to a paper by Black et al. (2005) which, along with earlier work by Dahl (2002), criticizes the standard practice of treating the returns to education as uniform across locations. They show that, in theory, the return to schooling is constant across locations only in the special case of homothetic preferences, and argue that the returns to education are empirically lower in high–amenity locations.[60]

Regional differences in income and poverty A related question arises when measuring earnings or income differences across locations. Using data for five US regions, Slesnick (2002) shows that regional comparisons of poverty rates based on nominal figures give a different picture than regional comparisons of poverty rates based on real figures. Consistent with the spatial equilibrium model, Slesnick shows that disparities in real income across regions are smaller when cost of living differentials are accounted for, so that the prevalence of poverty across regions changes significantly. For example, rural

---

[60] In a related paper, Black et al. (2009) argue that estimates of the wage differences between blacks and whites need to account for differences in the geographic location of different racial groups. They argue that accounting for geography changes the estimates of the speed of convergence between black and white earnings. They also develop a theoretical model to understand when estimates of the black–white earnings gap can be used to infer welfare differences.

areas and Southern states have a relatively low cost of living. As a consequence, poverty rates for rural areas and for many urban areas in the South are significantly closer to the rest of the country when real income is used instead of nominal income.

While this evidence is a useful first step, it is likely that an analysis at the regional level misses important intra-regional variation in the cost of living. A geographically more detailed analysis based on city-level data may uncover an even stronger effect of cost of living differences on the geographic distribution of poverty rates.

5.1.4. Subsidies to human capital when labor is mobile

In the spatial equilibrium model, workers' mobility determines the incidence of location-based government policies through their effects on land prices. Areas that benefit from government transfers experience an increase in the price of land that offsets (at least in part) the benefit of the transfer. However, in reality, workers' mobility may affect the incidence of location-based government policies even in the absence of local price changes. A salient example is the case of subsidies to human capital accumulation. Localized human capital externalities of the type discussed in Section 4.2.3 represent an important efficiency rationale for the provision of public subsidies to investment in human capital. However, in the presence of significant worker mobility it is not clear who ultimately benefits from these subsidies. Consider for example the case of subsidies to higher education. State and local government cover a larger fraction of higher education expenditures.[61] Yet, the high mobility of educated workers across states implies that part of the investment in human capital made by one state may benefit other states. Bound et al. (2004) quantify the magnitude of this problem by relating the production of new college graduates in a state to the stock of college-educated workers in the state. They find that the elasticity of stocks to flows is approximately 0.3 for BA's. This elasticity is even lower for students with medical degrees. This implies a high degree of migration. In a simple static model of where the supply and demand for college graduates determine the stock of college graduates, increases in the stock of college graduates due to increases in supply should lead to lower relative wages for college graduates, whereas increases in the stock due to increased demand should result in higher wages. The empirical results in Bound et al. (2004) are consistent with the prediction that inelastic local demand causes the effect of flows on stocks to be smaller. Additionally, a regression of relative wages on relative stocks indicates that the elasticity of wages to stocks is negative, consistent with some areas exporting college graduates and others importing them. Using more recent data, Bartik (2009) finds lower mobility rates of skilled workers and therefore concludes that state investment in higher education is not completely dissipated by labor mobility.

---

[61] For example, the current subsidy of direct costs to students at major public universities in the US is around 80%.

5.2. Efficiency considerations

The previous subsection concluded that the equity rationale for public transfers that target locations instead of individuals is generally not very compelling. A second possible rationale for government intervention has to do with efficiency. The key question in this respect is: are there market failures that suggest that governments should use taxpayer money to provide efficiency enhancing subsidies to firms to locate in their jurisdiction? In the absence of significant market failures, it is difficult to justify the use of taxpayer money for subsidies based on efficiency grounds. Here I consider four possible market failures.

5.2.1. Internalizing agglomeration spillovers

Local economic development policies are carried out both by local governments and by national governments.[62] To draw normative implications, it is therefore important to distinguish the point of view of a locality—which maximizes its own local welfare without consideration for aggregate welfare—from the point of view of the central government, which maximizes aggregate welfare.

Local welfare From the point of view of a local government, the most important efficiency rationale for location-based incentives is the existence of significant agglomeration externalities. This rationale hinges on whether the attraction of new businesses generates some form of external benefits to other firms in the same locality. If the attraction of new business generates localized positive agglomeration externalities, then the provision of subsidies may be able to generate the efficient allocation of resources in the local economy.[63]

In a static setting, the optimal magnitude of these incentives depends on the magnitude of agglomerations externalities. The literature described in Sections 4.1 and 4.2 suggests that these externalities may be important empirically, although there is still debate on their exact magnitude and the mechanisms that generate them.

In a dynamic setting, the existence of significant agglomeration economies has the potential to generate multiple equilibria. Kline (2010) proposes a simple theoretical framework that is useful in thinking about the magnitude of the efficient subsidy in the presence of multiple equilibria. In this case, the magnitude of the optimal subsidy is more complicated to derive, because it depends on the exact shape of Marshallian

---

[62] In the case of the European Union, some location-based policies are financed by the Union itself, which is an international government.

[63] An additional, although arguably less important, rationale that has been proposed for subsidies to attract new firms—and especially headquarters—has to do with charitable contributions. Card et al. (2007) document that attracting the headquarters of a publicly traded firm yields $3-10 million per year in contributions to local non-profits. Most of the increase in charitable contributions appears to be linked to the number of highly-compensated individuals in a city, rather than through direct donations by the corporations themselves. From a normative point of view, this rationale is not unassailable. Given the magnitude of the subsidies often required to attract headquarters, if this was the only benefit of attracting headquarters, there is no reason why the municipality should not have provided subsidies directly to the non-profit sector.

dynamics. Consider for example the case where the productivity of firms in a locality is a function of the number of other similar firms in that locality so that attracting new firms to a locality raises the productivity of all the firms in that locality (as in Greenstone et al., forthcoming). If the magnitude of this agglomeration externality is large enough, it is possible that the demand for labor is locally upward sloping, since more firms in a locality imply both more workers and higher productivity (as in Eq. (41) above). As discussed in Section 3.3, this setting is inherently characterized by multiple equilibria, with some equilibria featuring low levels of economic activity, low productivity and low nominal wages and other equilibria featuring high levels of economic activity, high productivity and high nominal wages. The size of the efficient subsidy can vary enormously, depending on the exact functional form of Marshallian dynamics, and on the starting point. On the one hand, if a locality is located at an unstable tipping point, a very limited subsidy can be enough to move it to a new equilibrium with an higher level of economic activity and agglomeration. On the other hand, if a locality is trapped in a stable bad equilibrium, "Big Push" type of policies may be needed to move it to a good equilibrium (Kline, 2010).

In this context, government intervention in the form of subsidies for firms to locate in the jurisdiction has the potential to start an agglomeration process that can ultimately shift a locality from a bad equilibrium (small agglomeration, low productivity) to a good equilibrium (large agglomeration, high productivity). The provision of a subsidy can have substantial and long lasting effects on the equilibrium level of economic activity in a locality.[64] Indeed, the expectation of government intervention alone may play a role in shifting a locality from a bad to a good equilibrium. In this case, the government policy acts as a coordination mechanism that signals to workers and firms which locality among all existing localities will move to a good equilibrium in the future. If firms and workers take this signal as credible, they will move to that jurisdiction, effectively realizing the expected outcome. In practice, it is implausible to expect that announcements alone are effective enough to have substantial real-world effects on the localization of economic activity. But it is plausible to think that in the presence of multiple equilibria, expectations together with the actual provision of subsidies might play a non-trivial role in the localization of firms and workers across locations.

In the presence of multiple equilibria, the efficiency benefits for location-based policies have the potential to be quite large relative to the efficiency costs. Consider the case where the provision of location based subsidies attracts new businesses to a location and move the locality from a bad equilibrium (low density of economic activity) to a good equilibrium (high density of economic activity). In this case the short-run efficiency costs of providing location-based subsidies could be small compared to the potential long-run benefits of moving to a better equilibrium.

[64] Of course, the process of agglomeration in this case would continue up to the point where the productivity advantages of agglomeration are offset by the increase in land prices (Section 3.3).

Aggregate welfare The efficiency argument, however, has different implications if one takes the point of view of the national government instead of the local government. For the point of view of a national government, the main concern is that the provision of subsidies by local governments may result in a zero-sum game, where the jobs created in targeted areas come at the expense of jobs elsewhere. In this context, justifying such policies requires a rationale for moving jobs from one location to another. In their comprehensive overview of the issue, Glaeser and Gottlieb (2009) argue that the only such rationale is for local agglomeration economies to be stronger on the margin in targeted areas. In other words, aggregate efficiency would require subsidies to favor areas that are more productive and where the elasticity of productivity with respect to agglomeration is higher. To achieve this efficient allocation, policy makers need to know the exact functional form of the spillover function. In practice, this functional form is still largely unknown, because it is difficult to estimate credibly. Given the difficulty of identifying the magnitude of agglomeration economies under a linearity assumption, it is not surprising that we still do not have a good idea of possible nonlinearities. Because of this difficulty, Glaeser and Gottleib conclude that policy-makers still do not have enough information to implement location-based policies that are efficient from the aggregate point of view.

However, when the benefits of attracting a new firm are highly heterogeneous, it is in principle possible that competition among localities may result in aggregate efficiency gains. Greenstone et al. (forthcoming) document one example of significant heterogeneity in productivity spillovers. Using the methodology described in Section 4.1, they present location-specific estimates of the impact on TFP of incumbent plants of new plant openings. Their Fig 2. reveals that there is substantial heterogeneity in the estimated spillover. Importantly, the magnitude of the productivity spillover does not appear to be random, but it varies systematically with easily observable features of the new plant. In particular, Greenstone, Hornbeck and Moretti document that the productivity spillover generated by the opening of a new large manufacturing plant is larger for incumbent plants that share similar labor and technology pools with the new plant.[65] It appears plausible in this case that policy makers are able to evaluate the potential value to their locality of attracting the new plant because this value depends on the degree of proximity of the local labor and technology pools to those of the new plant. This type of heterogeneity is important because it allows for the possibility of aggregate efficiency gains. National welfare is maximized when payments are made to plants that produce the spillovers so that they internalize this externality in making their location decision. In this case the locality that has the most to gain is the one that successfully attracts the new plant. The decentralized equilibrium in the presence of significant heterogeneity in

---

[65] By contrast, using a different methodology Glaeser and Gottlieb (2009) find little evidence that agglomeration economies vary systematically based on the easily observable characteristics of metropolitan areas. For example, they find that agglomeration economies do not appear to depend on city size or on urban amenities.

productivity spillovers together with the assumption that the potential value of attracting the new plant is known to each locality is not necessarily inconsistent with aggregate efficiency. From the equity point of view, the discussion of a locally financed subsidy in Section 5.1.1 suggests that in this case some but not all of the economic rent generated by the spillover ends up being transferred to the new firm. In particular, the locality that has the most to gain should only capture the fraction of the overall benefit that reflects the difference in production costs between the location with the most desirable attributes and the location with the second most desirable attributes.

5.2.2. *Unemployment, missing insurance and credit constraints*
Besides the existence of agglomeration externalities, there are three additional market failures that can in principle justify location-based policies.

First, wages may not be fully responsive to local shocks. Consider the case of idiosyncratic shocks to localities, where some cities or states are hit by negative demand shocks while other are not. The theoretical model in Section 3 is a full employment model, where wages always adjust fully to local shocks. The model indicates that if workers are mobile, and wages fully adjust, counter-cyclical transfers to localities hit by negative shocks will only have a limited effect on workers' welfare. However, the model effectively assumes away unemployment. In reality, wages may not be fully responsive to local labor market conditions, at least in the short run. If wages are not fully responsive, negative demand shocks will cause socially costly involuntary unemployment. In this case, countercyclical transfers from the central government have the potential to improve aggregate welfare. Of course, this potential welfare gain should be weighed against the distortionary cost of government intervention.

Second, even with full employment, homeowners are exposed to localized demand shocks, and there is no market-provided insurance to insure against such shocks. As explained in Section 3, shocks to a local economy affect the residents' value of housing, unless the elasticity of housing supply is infinite. Since housing is the most important asset for most households, the amount of risk generated by these shocks can be large. Existing financial instruments make it difficult to diversify housing risk. In this context, redistributive countercyclical policies that are centrally financed and target localities hit by negative idiosyncratic shocks may in principle act as government-provided insurance against housing value risk that the market does not insure. This policy effectively acts as a redistributive mechanism that transfers resources from homeowners in areas affected by positive shocks to homeowners in areas affected by negative shocks, thus reducing risk.

However, it is not clear that in practice these policies improve efficiency. First, there is the general deadweight loss of government intervention. Second, and more importantly, homeowners and workers often coincide. An optimal insurance scheme should therefore take into account both housing risk and wage risk. This is very difficult to do in practice because the correlation between housing values and wages is complicated. For example, the model with heterogenous labor in Section 3 has shown that if workers

have idiosyncratic preferences for location, the effect of a localized shock on a worker's welfare depends on that worker's skill level and location. If labor and housing supply elasticities are less then infinite, inframarginal skilled workers in the locality that receives a positive skill-biased productivity shock (city b) experience an increase in utility, while inframarginal unskilled workers experience a decrease. Inframarginal workers in city a, the city not directly affected by the shock, experience an increase in utility, irrespective of their skill level. In other words, in the example of the model, wages and housing values are negatively correlated for unskilled workers and for skilled workers in city a, but positively correlated for skilled workers in city b. This makes an efficient location-based insurance policy difficult to implement in practice.

Finally, the private and social costs of mobility might not be the same. When workers are mobile, and the private cost of mobility and the social cost of mobility are identical, the spatial adjustment that follows a negative shock to an area may be efficient. However, as pointed out by Blanchard and Katz (1992), credit constraints may introduce a wedge between the private and social costs of mobility. For example, following a negative shock to an area, workers without access to credit may be forced to leave even when it is optimal for them to borrow and wait for the equilibrium to be reestablished. This would lead to a socially inefficient degree of adjustment through mobility and may provide justification for efficiency-enhancing government intervention.

6. CONCLUSIONS

Understanding local labor markets is important for labor economists. The last three decades have witnessed a growing amount of empirical research on the causes and consequences of localization of workers and firms within a country. This area, at the intersection of labor and urban economics, contains important but challenging questions and is likely to generate an increasing share of high quality empirical research in the next decade.

Besides the obvious relevance for labor economists directly working on these topics, the idea of spatial equilibrium discussed in this chapter has a broader relevance for labor economists in general. It is difficult to understand the effect of nationwide labor demand changes on wages and employment without considering the role played by the spatial reallocation of labor and general equilibrium effects on local prices. Labor flows across localities and changes in local prices have the potential to undo some of the direct effects of labor market shocks.

The spatial equilibrium model presented in this chapter is a useful tool to think about the incidence of demand shocks when general equilibrium feedbacks are important. The version of the model that I propose in Section 3 is kept deliberately simple, so that all the equilibrium outcomes have transparent, closed-form solutions. Moreover, it is scalable, in the sense that it is relatively easy to relax some of the assumptions, in order to adapt it

to particular circumstances. The hope is that other researchers may find the framework useful in future work.

The survey of the empirical literature in Section 4 points to a growing body of solid empirical evidence on the existence of significant agglomeration economies. Although the econometric challenges—identification in particular—have proven in many cases difficult to fully overcome, there appears to be general agreement that agglomeration economies at the city level are empirically significant for many industries. While there is still debate on the exact economic magnitude of these externalities, many studies point to externalities of an economically non-trivial magnitude.

The last decade has also witnessed increasing efforts by researchers to pinpoint the precise mechanisms that might generate agglomeration economies. This is crucial to obtaining a convincing picture of the agglomeration phenomenon. Without understanding the precise mechanisms that generate agglomeration economies it is difficult to be confident about the existence of these externalities and to draw definitive conclusions for local development policies. Moreover, the three leading explanations imply different types of market failures and therefore call for different types of policy response. While the literature has produced a number of insightful empirical studies on the three possible mechanisms, overall the literature has not been completely successful in distinguishing between their relative importance. I share the view expressed by Glaeser and Gottlieb (2008) in a recent survey of this literature, when they conclude that "the field has still not reached a consensus on the relative importance of different sources of agglomeration economies". Given the important policy implications, more work is clearly needed on this topic.

Understanding the causes and the consequences of agglomeration of economic activity is crucial to understanding the economic rationale for location-based policies and their welfare consequences. These policies are widespread, but their economic rationales are not always clear. The discussion in Section 5 indicates that in a world where workers are mobile, targeting locations instead of individuals is an ineffective means of helping disadvantaged individuals. In a world with significant agglomeration spillovers, government intervention may be efficient from the point of view of a locality, although not always from the point of view of aggregate welfare.

UNCITED REFERENCES

Moretti, 2004c.

REFERENCES

Abdel Rahman, Hesham M., Fujita, Masahisa, 1990. Product variety, Marshallian externalities, and city sizes. Journal of Regional Science 30 (2), 165–183.
Acemoglu, Daron, 1996. A microfoundation for social increasing returns in human capital accumulation. Quarterly Journal of Economics 11 (August), 779–804.

Acemoglu, Daron, 1997. Training and innovation in imperfect labor markets. Review of Economic Studies.

Acemoglu, Daron, Angrist, Joshua, 2000. How large are human capital externalities? Evidence from compulsory schooling laws. NBER Macroeconomic Annual 15.

Adams, James D., Jaffe, Adam B., 1996. Bounding the effects of R&D: an investigation using matched establishment-firm data. RAND Journal of Economics, The RAND Corporation 27 (4), 700–721.

Albouy, David, 2009. The unequal geographic burden of federal taxation. Journal of Political Economy 117 (4), 635–667.

Andersson, Fredrik, Simon, Burgess, Lane, Julia I., 2007. Cities, matching and the productivity gains of agglomeration. Journal of Urban Economics 61 (1), 112–128.

Arzaghi, Mohammad, Henderson, J. Vernon, 2008. Networking off Madison Avenue. Review of Economic Studies 75 (4), 1011–1038.

Audretsch, David B., Feldman, Maryann P., 1996. R&D spillovers and the geography of innovation and production. American Economic Review 86 (3), 630–640.

Barro, R.J., Sala-i-Martin, X., 1991. Brookings Papers on Economic Activity.

Bartik, Tim, 1991. Who Benefits from State and Local Economic Development Policies? W.E. Upjohn Institute, Kalamazoo, MI.

Bartik, Timothy J., 2002. Evaluating the impacts of local economic development policies on local economic outcomes: what has been done and what is doable? Upjohn Working Papers 03-89, W.E. Upjohn Institute for Employment Research.

Bartik, Timothy J., 2009. What proportion of children stay in the same location as adults, and how does this vary across location and groups? Upjohn Working Papers 09-145, W.E. Upjohn Institute for Employment Research.

Baumgardner, James R., 1988. The division of labor, local markets, and worker organization. Journal of Political Economy 96 (3), 509–527.

Beaudry, Paul, Doms, Mark, Lewis, Ethan, 2006. Endogenous skill bias in technology adoption: city-level evidence from the IT revolution. NBER Working Papers 12521, National Bureau of Economic Research, Inc.

Bennmarker, Helge, Mellander, Erik, Öckert, Björn, 2009. Do Regional payroll tax reductions boost employment? Labour Economics 16, 480–489.

Berry, Christopher R., Glaeser, Edward L., 2005. The divergence of human capital levels across cities. NBER Working Papers 11617, National Bureau of Economic Research.

Black, Dan, Kolesnikova, N., Taylor, L., 2009. Earnings functions when wages and prices vary by location. Journal of Labor Economics 27 (1), 21–48.

Black, Dan, McKinnish, Terra, Sanders, Seth, 2005. The economic impact of the coal boom and bust. The Economic Journal 115 (April), 449–476.

Blanchard, Olivier, Katz, Larry, 1992. Regional evolutions. Brookings Papers on Economic Activity.

Bleakley, Hoyt, Lin, Jeffrey, 2006. Thick-market effects and churning in the labor market: evidence from US Cities, mimeo. University of Chicago.

Boarnet, Marlon, Bogart, William, 1996. Enterprise zones and employment: evidence from New Jersey. Journal of Urban Economics 40, 198–215.

Bondonio, Daniele, Engberg, John, 2000. Enterprise zones and local employment: evidence from the states' programs. Regional Science and Urban Economics 30 (5), 519–549.

Bound, John, Holzer, Harry J., 2000. Demand shifts, population adjustments, and labor market outcomes during the 1980s. Journal of Labor Economics 18 (1), 20–54.

Bound, John, Groen, Jeffrey, Kezdi, G. Gabor, Turner, Sarah, 2004. Trade in university training: cross-state variation in the production and stock of college-educated labor. Journal of Econometrics 121 (1–2), 143–173.

Busso, Matias, Gregory, Jesse, Kline, Patrick, 2009. Assessing the incidence and efficiency of a prominent place based policy, mimeo.

Card, David, Hallock, Kevin F., Moretti, Enrico, 2007. The geography of giving: the effect of corporate headquarters on local charities. University of California, Berkeley Mimeograph.

Carlino, Gerald A., Chatterjee, Satyajit, Hunt, Robert M., 2007. Urban density and the rate of invention. Journal of Urban Economics 61 (3), 389–419.

Carrington, William J., 1996. The Alaskan labor market during the pipeline era. Journal of Political Economy 104 (February).

Ciccone, Antonio, Hall, Robert E., 1996. Productivity and the density of economic activity. American Economic Review 86 (1), 54–70.

Combes, Pierre Philippe, Duranton, Gilles, Gobillon, Laurent, Puga, Diego, Roux, Sebastien, 2009. The productivity advantages of large cities: distinguishing agglomeration from firm selection. Discussion Paper 7191, Centre for Economic Policy Research.

Combes, Pierre Philippe, Duranton, Gilles, Overman, Henry G., 2005. Agglomeration and the adjustment of the spatial economy. Papers in Regional Science 84 (3), 311–349.

Costa, Dora L., Kahn, Matthew E., 2000. Power couples: changes in the locational choice of the college educated, 19401990. Quarterly Journal of Economics 115 (4), 1287–1315.

Dahl, Gordon, 2002. Mobility and the return to education: testing a Roy model with multiple markets. Econometrica 70 (6), 2367–2420.

David, Steve, Loungani, Prakash, Mahidhara, Ramamohan, 1997, Regional Labor Fluctuations: Oil Shocks, Military Spending, and other Driving Forces. Working paper, January.

Diamond, Peter A., 1982. Aggregate demand management in search equilibrium. Journal of Political Economy 90 (5), 881–894.

Doms, Mark, Lewis, Ethan, Robb, Alicia, 2009. Local labor market endowments, new business characteristics, and performance, mimeo, Federal Reserve Bank of San Francisco.

Duranton, Giles, Puga, Diego, 2004. Micro-foundations of urban agglomeration economies. In: Henderson, J. Vernon, Thisse, Jacques-François (Eds.), Handbook of Urban and Regional Economics, vol. 4. North-Holland, Amsterdam.

Duranton, Gilles, Overman, Henry G., 2005. Testing for localization using microgeographic data. Review of Economic Studies 72 (4), 1077–1106.

Duranton, Gilles, Puga, Diego, 2001. Nursery cities: urban diversity, process innovation, and the life cycle of products. American Economic Review 91 (5), 1454–1477.

Duranton, Gilles, Puga, Diego, 2004. Microfoundations of urban agglomeration economies. In: Henderson, J. Vernon, Thisse, Jacques-François (Eds.), Handbook of Regional and Urban Economics, vol. 4. NorthHolland, Amsterdam, pp. 2063–2117.

Eberts, Randall W., Stone, Joe A., 1992. Wage and Employment Adjustment in Local Labor Markets. Books from Upjohn Press, W.E. Upjohn Institute for Employment Research.

Ellison, Glenn, Glaeser, Edward L., 1997. Geographic concentration in US manufacturing industries: a dartboard approach. Journal of Political Economy 105 (5), 889–927.

Ellison, Glenn, Glaeser, Edward L., 1999. The geographic concentration of industry: does natural advantage explain agglomeration? American Economic Review Papers and Proceedings 89 (2), 311–316.

Ellison, Glenn, Glaeser, Edward L., Kerr, William, 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. American Economic Review (forthcoming).

Fallick, Bruce, Fleischman, Charles A., Rebitzer, James B., 2006. Job-hopping in silicon valley: some evidence concerning the microfoundations of a high-technology cluster. The Review of Economics and Statistics 88 (3), 472–481.

Faulk, Dagney, 2002. Do state economic development incentives create jobs? An analysis of state employment tax credits? National Tax Journal 55 (June), 263–280.

Glaeser, Edward L., 1998. Should transfer payments be indexed to local price levels? Regional Science and Urban Economics 28 (1), 1–20.

Glaeser, Edward L., 1999. Learning in cities. Journal of Urban Economics 46 (2), 254–277.

Glaeser, Edward L., Mare, David C., 2001. Cities and skills. Journal of Labor Economics 19 (2), 316–342.

Glaeser, Edward, Gottlieb, Joshua, 2009. The wealth of cities: agglomeration and the spatial equilibrium in the United States. Journal of Economic Literature.

Glaeser, Edward, Gottlieb, Joshua, 2008. The economics of place-making policies. Brookings Papers on Economic Activity.

Glaeser, Edward L., 2001. The economics of location-based tax incentives. Harvard Institute of Economic Research, Discussion Paper Number 1932.

Glaeser, Edward L., Kohlhase, Janet E., 2003. Cities, regions and the decline of transport costs. Papers in Regional Science 83, 197–228.

Glaeser, Edward, 2008. Cities, Agglomeration and Spatial Equilibrium. Oxford University Press.

Glaeser, Edward L., Scheinkman, Jose A., Shleifer, Andrei, 1995. Economic growth in a cross-section of cities. Journal of Monetary Economics 36 (1), 117–143.

Goolsbee, Austan, Maydew, Edward L., 2000. Coveting thy neighbor's manufacturing: the dilemma of state income apportionment. Journal of Public Economics 75, 125–143.

Greenstone, Michael, Moretti, Enrico, 2004. Bidding for industrial plants: does winning a "Million Dollar Plant" increase Welfare? National Bureau of Economic Research Working Paper No. 9844.

Greenstone, Michael, Hornbeck, Rick, Moretti, Enrico, Identifying Agglomeration spillovers: evidence from winners and losers of large plant openings, Journal of Political Economy (forthcoming).

Helsley, Robert W., Strange, William C., 1990. Matching and agglomeration economies in a system of cities. Regional Science and Urban Economics 20 (2), 189–212.

Henderson, J. Vernon, 2003. Marshall's scale economies. Journal of Urban Economics 55 (January), 1–28.

Henderson, J. Vernon, Black, Duncan, 1999. A theory of urban growth. Journal of Political Economy 107 (April), 252–284.

Holmes, Thomas J., 1999. Localization of industry and vertical disintegration. Review of Economics and Statistics 81 (2), 314–325.

Jaffe, Adam B., Trajtenberg, Manuel, Henderson, Rebecca M., 1993. Geographic localization of knowledge spillovers as evidenced by patent citation. Quarterly Journal of Economics 108 (August), 577–598.

Jovanovic, Boyan, Rob, Rafael, 1989. The growth and diffusion of knowledge. Review of Economic Studies 56 (October), 569–582.

Kline, Patrick, 2010. Place based policies, heterogeneity, and agglomeration, American Economic Review, Papers and Proceedings.

Krugman, Paul, 1991a. Geography and Trade. MIT Press, Cambridge, MA.

Krugman, Paul, 1991b. Increasing returns and economic geography. Journal of Political Economy 99 (june), 483–499.

Kuhn, Peter, Riddell, Chris, 2010. The long-term effects of a generous income support program: unemployment insurance in New Brunswick and Maine, 1940-1991, ILLR.

Lucas Jr., Robert E., 1988. On the mechanics of economic development. Journal of Monetary Economics 22 (July), 3–42.

Lucas Jr., Robert E., Rossi Hansberg, Esteban, 2002. On the internal structure of cities. Econometrica 70 (4), 1445–1476.

Machin, Steve, Pelkonen, Panu, Salvanes, Kjell G., 2009. Education and Mobility IZA WP 3845.

Malamud, Wozniak, 2009. The impact of college graduation on geographic mobility: identifying education using multiple components of Vietnam draft risk, IZA WP 3432.

Marshall, Alfred, 1890. Principles of Economics. Macmillan and Co., London.

Mas, Alex, Moretti, Enrico, 2009. Peers at work. American Economic Review 99 (1).

Melitz, Marc, Ottaviano, Gianmarco I.P., 2008. Market size, trade and productivity. Review of Economic Studies 75 (1), 295–316.

Mion, Giordano, Naticchioni, Paolo, 2009. The spatial sorting and matching of skills and firms. Canadian Journal of Economics, Canadian Economics Association 42 (1), 28–55.

Moretti, Enrico, 2004a. Estimating the external return to higher education: evidence from cross- sectional and longitudinal data. Journal of Econometrics 120 (July–August), 175–212.

Moretti, Enrico, 2004b. Workers' education, spillovers and productivity: evidence from plant- level production functions. American Economic Review 94 (June), 656–690.

Moretti, Enrico, 2004c. Human capital externalities in cities. In: Henderson, J. Vernon, Thisse, Jacques-François (Eds.), Handbook of Urban and Regional Economics, vol. 4. North-Holland, Amsterdam.

Moretti, Enrico, 'Local Multipliers' American Economic Review, Papers and Proceedings (forthcoming).

Notowidigdo, Matthew J., 2010. The Incidence of Local Labor Demand Shocks. MIT mimeo.

Overman, Henry G., Puga, Diego, 2010. Labour pooling as a source of agglomeration: an empirical investigation. In: Glaeser Edward L. (Ed.) Agglomeration Economics. Chicago University Press, Chicago, IL (forthcoming).

Papke, Leslie E., 1993. What do we know about enterprise zone? In: Poterba, J.M. (Ed.), Tax Policy and the Economy, vol. 7. MIT Press, Cambridge, Massachusetts.

Papke, Leslie E., 1994. Tax policy and urban development: evidence from the Indiana enterprise zone program. Journal of Public Economics 54, 37–49.

Partridge, Rickman, Li, 2009. Who wins from local economic development? Economic Development Quarterly 23 (1).

Peri, Giovanni, 2002. Young workers, learning, and agglomerations. Journal of Urban Economics 52 (3), 582–607.

Peters, Alan H., Fisher, Peter S., 2002. State enterprise zone programs: have they worked? W.E. Upjohn Institute, Kalamazoo, MI.

Petrongolo, Barbara, Pissarides, Christopher A., 2005. Scale effects in markets with search. The Economic Journal 116 (January), 21–44.

Puga, Diego, 2009. The magnitude and causes of agglomeration economies. Mimeo.

Rauch, James E., 1993. Productivity gains from geographic concentration of human capital: evidence from the cities. Journal of Urban Economics 34 (November), 380–400.

Renkow, M., 2003. Employment growth, worker mobility and rural economic development. American Journal of Agricultural Economics 85 (2), 503–513.

Renkow, M., 2006. Employment growth and the allocation of new jobs. Review of Regional Studies 36 (1), 121–139.

Roback, Jennifer, 1982. Wages, rents and the quality of life. Journal of Political Economy 90 (December), 1257–1278.

Roback, Jennifer, 1988. Wages, rents and amenities: differences among workers and regions. Economic Inquiry 26 (January), 23–41.

Rosen, Sherwin, 1979. Wagebased indexes of urban quality of life. In: Miezkowski, Peter N., Straszheim, Mahlon R. (Eds.), Current Issues in Urban Economics. Johns Hopkins University Press, Baltimore, MD, pp. 74–104.

Rosenthal, Stuart S., Strange, William C., 2001. The determinants of agglomeration. Journal of Urban Economics 50 (September), 191–229.

Rosenthal, Stuart S., Strange, William C., 2004. Evidence on the nature and sources of agglomeration. In: Henderson, J. Vernon, Thisse, Jacques-François (Eds.), Handbook of Urban and Regional Economics, vol. 4. North-Holland, Amsterdam.

Rosenthal, Stuart S., Strange, William C., 2003. Geography, industrial organization, and agglomeration. Review of Economics and Statistics 85 (2), 377–393.

Rosenthal, Stuart S., Strange, William C., 2008. The attenuation of human capital spillovers. Journal of Urban Economics 64 (2), 373–389.

Rotemberg, Julio R., Saloner, Garth, 2000. Competition and human capital accumulation: a theory of interregional specialization and trade. Regional Science and Urban Economics 30 (4), 373–404. doi:10.1016/S0166-0462(99)00044-7. (ISSN: 0166-0462).

Saxenian, Anna Lee., 1994. Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, MA.

Slesnick, Daniel T., 2002. Prices and regional variation in welfare. Journal of Urban Economics 51, 446–468.

Sveikauskas, Leo, 1975. Productivity of cities. Quarterly Journal of Economics 89 (3), 393–413.

Tabuchi, Takatoshi, Thisse, Jacques-François, 2002. Taste heterogeneity, labor mobility and economic geography. Journal of Development Economics 69 (1), 155–177.

Topel, Robert H., 1986. Local labor markets. The Journal of Political Economy 94 (3).

Wasylenko, Michael, 1997. Taxation and economic development: the state of the economic literature. New England Economic Review (March/April), 37–52.

Wheaton, William C., Lewis, Mark J., 2002. Urban wages and labor market agglomeration. Journal of Urban Economics 51 (3), 542–562.

Wheeler, Christopher, 2008. Local market scale and the pattern of job changes among young men. Federal Reserve Bank of St Louis Working Paper 2005-033C. Revised.

Zucker, Lynne G., Darby, Michael R., Brewer, Marilynn B., 1998. Intellectual human capital and the birth of US biotechnology enterprises. American Economic Review 88 (March), 290–306.