

NBER WORKING PAPER SERIES

WHY DO CITIES MATTER? LOCAL GROWTH AND AGGREGATE GROWTH

Chang-Tai Hsieh
Enrico Moretti

Working Paper 21154
<http://www.nber.org/papers/w21154>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2015

We are grateful to Klaus Desmet, Rebecca Diamond, Daniel Fetter, Cecile Gaubert, Ed Glaeser, Eric Hurst, Pat Kline, Steve Redding and seminar participants at the AEA Meetings, Brown, Chicago Booth, Chicago Fed, Minneapolis Fed, NBER Summer Institute, Simon Fraser University, Stanford SIEPR, University of Arizona, Università Cattolica di Milano, University of British Columbia, UC Berkeley Haas, UCLA Anderson, University of Venice, and Yrjo Jahnsson Foundation for useful suggestions. This paper previously circulated under the title "Growth in Cities and Countries." The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Chang-Tai Hsieh and Enrico Moretti. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Why Do Cities Matter? Local Growth and Aggregate Growth
Chang-Tai Hsieh and Enrico Moretti
NBER Working Paper No. 21154
May 2015, Revised June 2015
JEL No. E24,J01,R0

ABSTRACT

We study how growth of cities determines the growth of nations. Using a spatial equilibrium model and data on 220 US metropolitan areas from 1964 to 2009, we first estimate the contribution of each U.S. city to national GDP growth. We show that the contribution of a city to aggregate growth can differ significantly from what one might naively infer from the growth of the city's GDP. Despite some of the strongest rate of local growth, New York, San Francisco and San Jose were only responsible for a small fraction of U.S. growth in this period. By contrast, almost half of aggregate US growth was driven by growth of cities in the South. We then provide a normative analysis of potential growth. We show that the dispersion of the conditional average nominal wage across US cities doubled, indicating that worker productivity is increasingly different across cities. We calculate that this increased wage dispersion lowered aggregate U.S. GDP by 13.5%. Most of the loss was likely caused by increased constraints to housing supply in high productivity cities like New York, San Francisco and San Jose. Lowering regulatory constraints in these cities to the level of the median city would expand their work force and increase U.S. GDP by 9.5%. We conclude that the aggregate gains in output and welfare from spatial reallocation of labor are likely to be substantial in the U.S., and that a major impediment to a more efficient spatial allocation of labor are housing supply constraints. These constraints limit the number of US workers who have access to the most productive of American cities. In general equilibrium, this lowers income and welfare of all US workers.

Chang-Tai Hsieh
Booth School of Business
University of Chicago
5807 S Woodlawn Ave
Chicago, IL 60637
and NBER
chsieh@chicagoBooth.edu

Enrico Moretti
University of California, Berkeley
Department of Economics
549 Evans Hall
Berkeley, CA 94720-3880
and NBER
moretti@econ.berkeley.edu

1. Introduction

Macroeconomists have long been fascinated by the vast differences in economic activity between countries. Yet, differences between cities or regions within each country are equally striking. While there is a long tradition of using cities or regions as laboratories to understand the sources of differences across countries (for example, Barro and Sala-i-Martin, 1991 and 1992; Gennaioli, LaPorta, Lopez-de-Silanes, and Shleifer, 2013), researchers have paid less attention to how the geographical distribution of economic activity across cities or regions itself affects aggregate outcomes of a given country. At the same time, a large urban economics literature identifies local forces that explain differences in wages and economic activity across cities but has paid comparatively less attention to how these forces aggregate to affect growth for the country as a whole.

In this paper we bridge this gap. We study how economic growth of cities determines the growth of nations. We use data on 220 US cities over the past five decades and a spatial equilibrium model to address two related questions---a positive one and a normative one. First, we estimate the contribution of each US metropolitan area to aggregate output growth between 1964 and 2009. We show that our model-based calculation of a given city's contribution to aggregate growth differs significantly from what one might naively infer from the growth of the city's GDP. We then turn to a *normative* analysis of potential growth. We document a significant increase in the spatial dispersion of wages between 1964 and 2009, indicating that worker productivity is increasingly different across American cities. We argue that these productivity differences reflect an increasingly inefficient spatial allocation of labor across US cities, and that much of this inefficiency is caused by restrictive housing policies of municipalities with high productivity, like New York and San Francisco.

We base our analysis on a Rosen-Roback model where workers can freely move across cities and geographical differences in wages reflect differences in local labor demand and supply. In turn, local labor demand reflects forces that affect the TFP of firms in a city---infrastructure, industry mix, agglomeration economies, human capital spillovers, access to non-tradable inputs and local entrepreneurship---while local labor supply reflects amenities and housing supply.

We analyze how these local forces aggregate in the Rosen-Roback model to affect national output and welfare. We show that aggregate output increases in local TFP in each city but decreases in the dispersion of wages across cities.¹ The reason is that wage dispersion across

¹ Formally, we show that aggregate output growth can be decomposed into the contribution of the weighted average of the growth rate of local TFP and into the change in the dispersion of the marginal product of labor across cities.

cities reflects variation in the marginal product of labor: The wider the dispersion of marginal products across cities, the lower aggregate output, everything else constant. Intuitively, if labor is more productive in some areas than in others, then aggregate output may be increased by reallocating some workers from low productivity areas to high productivity ones.

In this setting, geography matters in the sense that the same localized shock can have profoundly different aggregate effects depending on where it takes place. An increase in local labor demand caused by a TFP increase will have a large effect on aggregate output if the TFP increase generates an increase in local employment, but the same increase in local TFP in another city can have a much smaller aggregate effect if it largely results in higher nominal wages in the city and only a small increase in local employment.

Empirically, we begin by calculating the contribution of each US city to aggregate growth and compare it with an accounting measure based solely on the growth of the city's GDP. We show there are large differences between these two numbers. For example, growth of New York's GDP was 12 percent of aggregate output growth from 1964 to 2009. However, viewed through the lenses of the Rosen-Roback model, New York was only responsible for less than 5 percent of aggregate output growth. The difference is because much of the output growth in New York was manifested as higher nominal wages, which increased the overall spatial misallocation of labor. On the other extreme, Detroit's GDP fell dramatically from 1964 to 2009, but its net contribution to aggregate output growth was actually positive. In the case of Detroit, the decline in its nominal wage from 1964 to 2009 lowered the overall wage dispersion.

We then turn from a positive analysis of the local forces underlying aggregate growth to a normative analysis of potential growth. We focus on the effects of the growing dispersion in the marginal product of labor across cities. We show that after conditioning on workers characteristics, the geographical distribution of nominal wages is significantly wider today than in 1964. In particular, the standard deviation of conditional wages across US cities in 2009 is twice as large compared to 1964, indicating that differences in worker productivity across cities are growing.

When we quantify the output and welfare cost of this increase in dispersion of the marginal product of labor, we find that aggregate output in 2009 would have been significantly higher if the dispersion of nominal wages had not increased. Holding the distribution of local TFP fixed at 2009 levels, we hypothetically reallocate labor from high wage to low wage cities such that the

The growth rate of aggregate output is higher when the weighted average of the growth rate of local TFP is higher and is lower when the weighted average of wage dispersion across cities increases (for a given distribution of TFP).

hypothetical wage in each city (relative to the average wage) is equal to the relative wage in 1964. Intuitively, this scenario involves setting amenities and housing supply at their 1964 level, while keeping labor demand constant at its 2009 level, and allowing workers to reallocate across cities in response. Under this scenario, aggregate yearly GDP growth from 1964 to 2009 would have been 0.3 percentage points higher. In levels, U.S. GDP in 2009 would be 13.5% or \$1.95 trillion higher. This amounts to an annual wage increase of \$8775 for the average worker.

This output effect is driven to a large extent by three cities -- New York, San Francisco and San Jose – which experienced some of the strongest growth in labor demand over the last four decades, thanks to growth of human capital intensive industries like high tech and finance (Moretti, 2012). But most of the labor demand increase was manifested as higher nominal wages instead of higher employment. The resulting increase in overall wage dispersion negatively impacted aggregate growth. In contrast, Southern cities also experienced rapid output growth, but much of this growth showed up as employment growth and only a small amount as an increase in the nominal wage. The resulting decrease in overall wage dispersion fostered aggregate growth, although the impact was smaller than that one in New York, San Francisco and San Jose.

Of course, the potential output gains from spatial reallocation of labor do not necessarily translate into welfare gains. The effect on aggregate welfare depends on why wages are not equalized across cities in the first place.² If the relative increase in nominal wages in high TFP cities such as San Francisco and New York is due to restrictions to housing supply, then the aggregate output loss due to differences in the marginal product of labor also imply welfare losses. In this case, removing constraint to housing supply in cities like San Francisco and New York would allow more workers to move there and take advantage of their higher productivity, increasing both aggregate output and welfare. In contrast, if labor supply in New York and San Francisco is low because of increasingly undesirable local amenities, then the loss in aggregate output from the gaps in the marginal product of labor does not necessarily reflect a loss in welfare. For example, if equilibrium wages in New York are high because people dislike congestion, noise and pollution and need to be compensated for it, then moving more people to New York will increase aggregate output, but will lower welfare.

² Formally, we show that aggregate welfare in the Rosen-Roback model is simply aggregate output divided by a weighted average of the ratio of local housing prices to local amenities. Holding aggregate output constant, higher housing prices lower aggregate welfare and better local amenities increase welfare.

When we decompose the increase in wage dispersion into the changes due to housing supply and amenities, we find that the increase is almost entirely driven by the former. Setting amenities back to their 1964 levels slightly decreases the overall wage dispersion and increases aggregate output, but the effect is quantitatively small. In contrast, we find that constraints to housing supply in cities with high TFP are a major driver of our findings. We use data from Saiz (2010) to separate overall elasticity of housing supply in each U.S. city into the availability of land and municipal regulations.

We estimate that holding constant land but lowering land use regulations in New York, San Francisco and San Jose to the level of the median city would increase U.S. output by 9.7%. In essence, more housing supply would allow more American workers to access the high productivity of these high TFP cities. We also estimate that increasing regulations in the South would be costly for aggregate output. In particular, we estimate that increasing land use regulations in the South to the level of New York, San Francisco and San Jose would lower U.S. output by 3%.

We conclude that the aggregate gains in output and in welfare from spatial reallocation of labor are likely to be substantial in the U.S., and that a major impediment to a more efficient spatial allocation of labor is the growing constraints to housing supply in high wage cities. These constraints limit the number of US workers who can work in the most productive of American cities. In general equilibrium, this lowers income and welfare of *all* US workers and amount to a large *negative externality* imposed by a minority of cities on the entire country.

This paper builds on two bodies of work. First, we build on the large empirical work, beginning with Rosen (1979) and Roback (1982), on local labor supply and labor demand. The effect of stringent land use regulations on *local* housing prices is well documented (Glaeser, Gyourko and Saks, 2005 and 2006; Gyourko and Glaeser, 2005; Saiz, 2010), and our paper highlights the *aggregate* negative impacts of such regulations (and the positive effect of the relative absence of such regulations in the US South). Our findings on the importance of housing supply constrains are consistent with those in Ganong and Shoag (2013). Second, we draw on the theoretical work on systems of cities in spatial equilibrium. In particular, Henderson (1981, 1982), Au and Henderson (2006a and 2006b), Behrens et. al. (2014), Eeckout et. al. (2014), Desmets and Rossi-Hansberg (2013) and Redding (2014) model the equilibrium allocation of resources across cities. Our approach is most closely related to Desmets and Rossi-Hansberg (2013), Redding (2014) and Gaubert (2014). Desmets and Rossi-Hansberg (2013) analyze the effect on the heterogeneity of local TFP, amenities and “local frictions” in the US and China,

Redding (2014) analyzes on the effect of internal trade frictions, and Gaubert (2014) analyzes optimal city size. We abstract from trade frictions and heterogeneity in local TFP to focus on the effect of local housing supply on wage dispersion, aggregate output and welfare. Another closely related paper is Duranton et al. (2015) who quantify the misallocation of manufacturing output in India caused by misallocation of land. Our finding of barriers to labor mobility in the U.S. complements the finding of broader set of barriers to factor mobility in 83 countries in Gennaioli, LaPorta, Lopez-de-Silanes, and Shleifer (2014).

The paper is organized as follows. In Section 2 we present the model. In Section 3 we describe the data and the key changes in wage dispersion. Empirical findings are in Section 4. Section 5 discusses policy implications.

2. Model

This section examines the channels by which local forces in a city affect aggregate output and welfare. The model is a standard Rosen-Roback model with a spatial equilibrium. Cities differ by local labor demand and local labor supply. Specifically, city i produces a traded good sold at a fixed price in the national market with the following technology

$$(1.1) \quad Y_i = A_i L_i^\alpha K_i^\eta.$$

Here, A_i denotes total factor productivity, L_i employment, and K_i capital. We assume $\alpha + \eta < 1$. We interpret A_i as capturing forces such as cost advantages enjoyed by firms in the city (access to waterways, railways, airports, topography, nature of the terrain, weather, local institutions, labor and environmental regulations), demand for products made by the city, ease of entry, agglomeration economies or technological spillovers that benefit all firms in the city.

Workers can freely move across cities and their indirect utility given by

$$(1.2) \quad V = \frac{W_i Z_i}{P_i^\beta}.$$

Here W_i denotes the nominal wage, Z_i amenities, P_i the price of housing in city, and β is the share of expenditures on housing.³ We assume that capital is supplied with infinite elasticity at an exogenously given rental price.

³ While different cities have different income and different prices, the share of expenditures on housing does not vary with income (Davis and Ortalo-Magnes, 2010; Lewbel, Arthur and Krishna Pendakur, 2008), which suggests that β is roughly constant.

We make several simplifying assumptions. First, the expression for indirect utility implicitly assumes workers do not own the housing stock, but rent from an absentee landlord. Second, we assume that workers have homogeneous tastes over locations and are perfectly mobile across locations. This makes labor supply to a local labor market infinitely elastic. Third we assume that TFP and amenities can vary across cities but are exogenous. Fourth, we assume all cities produce the same product and do not specialize. Finally, we assume no heterogeneity in labor demand elasticity. We relax all these assumptions later.

We now solve for the equilibrium allocation of employment and wage across cities. First, equating the marginal product of labor to the cost of labor in each city and the cost of capital to an exogenously determined interest rate, employment is:

$$(1.3) \quad L_i \propto \left(\frac{A_i}{W_i^{1-\eta}} \right)^{\frac{1}{1-\alpha-\eta}}$$

Employment is increasing in local TFP and decreasing in the nominal wage, with an elasticity that depends on the slope of the labor demand curve. After substituting (1.2) into (1.3)

employment can also be expressed as $L_i \propto \left(\frac{A_i Z_i^{1-\eta}}{P_i^{\beta(1-\eta)}} \right)^{\frac{1}{1-\alpha-\eta}}$. Not surprisingly, cities with more employment are those with high local TFP, low housing prices, or high quality amenities.

We assume housing prices reflect local demand and supply conditions. Specifically, we assume $P_i = L_i^{\gamma_i}$ where γ_i is a parameter that governs the elasticity of housing supply with respect to the number of workers. An increase in the number of workers has a larger effect on housing prices when γ_i is large. We think of heterogeneity in γ_i as capturing differences in both land availability and housing regulations (such as land use regulations). Cities with limited amount of land and stringent land use regulations have a large γ_i ; cities with abundant land and permissive land use regulations have a small γ_i (Glaeser, Gyourko and Saks, 2005 and 2006; Saiz, 2010).

We can now write the equilibrium wage as a function of three exogenous factors:

$$(1.4) \quad W_i \propto \left(\frac{A_i^{\beta\gamma_i}}{Z_i^{1-\alpha-\eta}} \right)^{\frac{1}{(1-\eta)(1+\beta\gamma_i)-\alpha}}$$

The equilibrium wage is increasing in local TFP and decreasing in amenities with an elasticity that depends on the local elasticity of housing supply γ_i . The first factor – local TFP – reflects

labor demand. Higher local TFP implies stronger demand for labor and therefore higher equilibrium nominal wages, *ceteris paribus*. The other two factors -- amenities and housing supply reflect labor supply. In equilibrium, better amenities imply larger supply and therefore lower nominal wages. Intuitively, the utility stemming from the amenities makes workers willing to live in a city even if their nominal wages are lower. More elastic housing supply also implies lower wages, but for a different reason. More elastic labor supply means that in cities with positive demand or amenity shocks, the cost of housing increase by less. The spatial variation of wages reflects the variation in TFP, amenities, and housing supply and the covariance between these variables.

2.1 Aggregate Output and Welfare

We now solve for aggregate output and welfare. First, we use (1.2) and (1.3) to express welfare as:

$$(1.5) \quad V = Y \cdot \left(\sum_i L_i \cdot \frac{P_i^\beta}{Z_i} \right)^{-1}$$

where $Y = \sum_i Y_i$ denotes aggregate output. Intuitively, welfare is aggregate output in units of utility and $\sum_i L_i \cdot \frac{P_i^\beta}{Z_i}$ is the cost minimizing price of a unit of utility (the price of goods is normalized to one).⁴ Second, we solve for aggregate output by imposing the condition that aggregate labor demand is equal to aggregate labor supply (normalized to one):

$$(1.6) \quad Y = \sum_i Y_i \propto \left(\sum_i A_i \frac{1}{W_i^{1-\alpha-\eta}} \left(\frac{\bar{W}}{W_i} \right)^{\frac{1-\eta}{1-\alpha-\eta}} \right)^{\frac{1-\alpha-\eta}{1-\eta}}$$

where $\bar{W} \equiv \sum_i W_i \cdot L_i$ denotes the employment-weighted average nominal wage and W_i is determined by (1.4). Aggregate output is a harmonic mean of the product of local TFP and the inverse of the wage gap of the city relative to the mean wage. Housing supply restrictions affect

⁴ Equation (1.5) only considers the effect of labor income on welfare. If we instead assume that firm profits accrue to the workers, the sum of labor income and profits is proportional to aggregate output. Therefore, welfare would still be proportional to aggregate output divided by the price of utility.

welfare through their effect on the average price of housing and on aggregate output by changing the dispersion of nominal wages.

We can now decompose the sources of aggregate growth in output and welfare. The growth of aggregate output is:

$$(1.7) \quad \frac{Y_{t+1}}{Y_t} = \left(\frac{\sum_i A_{i,t+1}^{\frac{1}{1-\alpha-\eta}}}{\sum_i A_{i,t}^{\frac{1}{1-\alpha-\eta}}} \right)^{\frac{1-\alpha-\eta}{1-\eta}} \left(\frac{\sum_i \tilde{L}_{i,t+1} \cdot \left(\frac{\bar{W}_{t+1}}{W_{i,t+1}} \right)^{\frac{1-\eta}{1-\alpha-\eta}}}{\sum_i \tilde{L}_{i,t} \cdot \left(\frac{\bar{W}_t}{W_{i,t}} \right)^{\frac{1-\eta}{1-\alpha-\eta}}} \right)^{\frac{1-\alpha-\eta}{1-\eta}}$$

where $\tilde{L}_i \equiv \left(A_i / \left(\sum_j A_j^{\frac{1}{1-\alpha-\eta}} \right)^{1-\alpha-\eta} \right)^{\frac{1}{1-\alpha-\eta}}$ denotes the hypothetical city size when wages are the same in all cities.

Equation (1.7) suggests that aggregate output growth can be decomposed into the effect of local TFP (the first term in (1.7)) and into the effect of changes in the spatial dispersion of wages

(the second term in (1.7)). The effect of spatial dispersion is given by $\sum_i \tilde{L}_i \cdot \left(\frac{\bar{W}}{W_i} \right)^{\frac{1-\eta}{1-\alpha-\eta}}$ measured

in the two years. Intuitively, this term measures the ratio of aggregate output observed in each year to the hypothetical output when wages were the same in all cities in that year (and labor and capital is reallocated in response to the change in the wage distribution). Because the exponent on the relative wage is greater than one, aggregate output rises when wage dispersion falls (holding local TFP fixed).⁵

The growth of aggregate welfare depends on the same two forces as well as on changes in the price of utility, because we have seen in equation (1.5) that aggregate welfare is equal to aggregate output times the price of utility (i.e. the weighted average of the ratio of local amenities to the local housing price.) Thus there are three channels via which local shocks affect aggregate welfare: the price of utility, the weighted average of local TFP, and the weighted dispersion of wages across cities.

To illustrate these mechanisms, consider how changes in local TFP or local amenities affect aggregate output and welfare. First, suppose that local TFP rises in a city. This raises the

⁵ We assume decreasing returns to scale. With constant or increasing returns to scale, the distribution of employment would be degenerate as the city with the highest TFP would attract all economic activity.

weighted average of local TFP, which increases aggregate output and welfare (holding the price of utility constant). The increase in local TFP also raises the local housing price by increasing the local demand for housing. This increases the price of utility in *all* cities, which lowers welfare (holding aggregate output fixed), and this effect is larger when the local housing supply is inelastic. Finally, high local housing prices increases the local wage, but the aggregate effect of a higher local wage is ambiguous. If the high local housing price increases the gap between the local wage and the average wage, aggregate output -- and welfare -- falls. When this is the case, the growth rate of local GDP *overstates* the contribution of the local growth to aggregate output growth. But if the TFP increase occurs in a low wage city, the increase in the local wage potentially lowers the overall wage dispersion, which increases aggregate output. In this case, the growth rate of local GDP *understates* the local contribution to aggregate GDP.

Second, consider the effect of a decline in local TFP. Low TFP lowers the average of local TFP, which lowers aggregate output and welfare. In addition, Glaeser and Gyourko (2005) show that housing prices drop sharply in cities that suffer from an adverse labor demand shock. In our framework, the decline in housing prices has two additional effects. First, lower housing prices lowers the price of utility, which offsets the effect of lower aggregate output on welfare. The drop in housing prices also lowers the local nominal wage, but as before the aggregate effect depends on whether the local nominal wage was above or below the nationwide mean. If the local wage is above the mean, the decline in the nominal wage potentially narrows the marginal product gap, which increases aggregate output and welfare. In this case, local GDP falls because of the direct effect of the decline in local TFP and the fall in the local wage. However, even when local output growth is negative, the net effect on aggregate output growth may well be positive if the effect of the narrowing wage dispersion is larger than the direct effect of the decline in local TFP. In the empirical results, we will show that this appears to have been the case in many US cities where local TFP fell.

Third, consider the effect of an improvement in amenities. When amenities improve in high wage cities, this increases the average level of amenities and lowers overall wage dispersion. Here the decline in wage dispersion unambiguously improves welfare, and the local output growth understates the contribution of the local economy to aggregate output. On other hand, when amenities improve in low wage cities, this also increases the average level of amenities, but increases the overall wage dispersion. Here, although local GDP increases, the improvement in amenities *lowers* aggregate output. The output decline due to increased wage dispersion offsets some of the direct effect of the improvement in the average level of amenities.

In the empirical section of the paper, we use this framework to provide two calculations. First, we measure the contribution of each US city to aggregate US growth. We show that the model-based calculation of the contribution of each city to aggregate growth is empirically quite different from a naïve accounting-based calculation based on the measured growth of local output. Second, we use this framework to calculate the counterfactual output and welfare growth in the US under different assumptions on wage dispersion. We ask how much faster output and welfare growth would have been if wage dispersion had not increased in the US but had remained constant and link the increase in wage dispersion to specific housing supply policies on the part of US cities.

2.2 Extensions

We now consider the effect of several extensions of our basic model.

Ownership of Housing Stock: We have assumed that workers do not own the housing stock so that an increase in average housing prices lowers welfare holding aggregate output fixed. Suppose we assume instead that the housing stock is owned by the workers in equal proportions, irrespective of where they live. Think of workers as owning equal shares in a mutual fund that own all the housing in the US. All the equations are the same, except that welfare is given by

$$V \propto \left(Y + \sum_i L_i h_i P_i \right) \cdot \left(\sum_i L_i \cdot \frac{P_i^\beta}{Z_i} \right)^{-1} \quad \text{where } h_i \text{ denotes per-capita housing consumption in city } i.$$

After imposing the condition that the share of nominal expenditures on housing is equal to β , the change in housing prices has the same effect on nominal income as on the average price of housing. In this case, changes in housing prices only affect welfare through the effect of the dispersion of the nominal wage on aggregate output, but changes in the *average* price of housing has no effect on welfare.

The most realistic case is of course the one where workers own housing in the city where they live. In this case, changes in house prices induced by our counterfactuals have redistributive effects: workers in some areas are made better off, while workers in other areas are made worse off. But *in the aggregate*, the conclusions are identical to the case in which the housing stock is owned by the workers in equal proportions, irrespective of where they live: housing prices only affect welfare through the effect of the dispersion of the nominal wage on aggregate output. Thus, estimates of the effects based on the baseline model remains valid in the aggregate.

Specialization by Cities: Our baseline model assumes that the output of a city is a perfect substitute for the products made by other cities. Suppose instead that each city makes a differentiated product with a production function given by $Y_i = A_i L_i$. The demand for the

product of each city is determined by utility defined as $U_j = \left(\sum_i Y_{ij}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma(1-\beta)}{\sigma-1}} h_j^\beta Z_j$ where U_j

denotes utility in city j , Y_{ij} denotes consumption of city i 's output in city j , and h_j is per-capita

housing in city j . The labor demand in each city is given by $L_i \propto \left(\frac{A_i}{W_i} \right)^{\sigma-1}$ and aggregate output

by $Y = \left(\sum_i A_i^{\sigma-1} \left(\frac{\bar{W}}{W_i} \right)^{\sigma-1} \right)^{\frac{1}{\sigma-1}}$. These last two equations are identical to (1.3) and (1.6) when we

substitute $\frac{1}{1-\alpha}$ with $\sigma-1$.⁶ In words, a model with constant returns to scale and where cities

are specialized in production is isomorphic to a model where cities produce identical products and with a decreasing returns to scale production function. Finally, assuming that the output

good is available in all cities at the same price, we can normalize the cost-minimizing price of

one unit of the CES aggregate of the output good $\left(\sum_i Y_i^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$ to one. With this normalization,

welfare is still given by (1.5).

Imperfect Labor Mobility: We can also relax the assumption of infinite labor mobility.

Suppose that workers differ in preferences over locations. Specifically, suppose the indirect

utility of worker j in city i is given by $V_{ji} = \varepsilon_{ji} \cdot \frac{W_i Z_i}{P_i^\beta}$ where ε_{ji} is a random variable measuring the

taste of individual j in city i as, for example, in Moretti (2010). A larger ε_{ji} means that worker i

is particularly attached to city j for idiosyncratic reasons. We assume that workers locate in the

city where her utility V_{ji} is maximized. In this case, workers tend to move toward cities with

high real wages and good amenities, but they are not infinitely sensitive to small wage

⁶ Since labor is the only factor of production in the differentiated products model, we set the capital share to zero in the baseline model for comparability.

differences. The implication is that only marginal workers are indifferent across cities and all the other workers are infra-marginal.

To make this model tractable, we assume that ε_{jt} are independently distributed and drawn from a multivariate extreme value distribution. Specifically, we follow Kline and Moretti (2013) and assume the joint distribution of ε_{jt} is given by $F_g(\varepsilon_1, \dots, \varepsilon_N) = \exp\left(-\sum_i^N \varepsilon_i^{-\theta}\right)$ where the parameter $1/\theta$ governs the strength of idiosyncratic preferences for location and therefore the degree of labor mobility. If $1/\theta$ is large, many workers require large real wage or amenity differences to be compelled to move. On the other hand, if $1/\theta$ is small, most workers are not particularly attached to one community and will be willing to move in response to small differences in real wages or amenities.⁷

In this model, employment in a city is still given by (1.3) and aggregate output by (1.6). What is new is that the Rosen-Roback condition that differences in wages across cities are directly proportional to the ratio of housing prices to amenities (equation (1.5)) no longer holds. Instead, the (inverse) labor supply equation of a city is given by:

$$(1.8) \quad W_i \propto \frac{P_i^\beta L_i^{\frac{1}{\theta}}}{Z_i}$$

This says even when housing prices and amenities are the same in all cities, wages will differ between large and small cities with an elasticity that depends on the heterogeneity in preferences for location. Intuitively, higher wages in large cities are needed to compensate marginal individuals to live in the city. When we endogenize the housing price as a function of city size and the local housing supply elasticity and impose the condition that labor demand is equal to labor supply, the equilibrium nominal wage is given by:

$$(1.9) \quad W_i \propto \left(\frac{1}{Z_i}\right)^{\frac{1-\alpha-\eta}{(1-\alpha-\eta)+(1-\eta)(\beta\gamma_i+1/\theta)}} A^{\frac{\beta\gamma_i+1/\theta}{(1-\alpha-\eta)+(1-\eta)(\beta\gamma_i+1/\theta)}}$$

Finally, while utility differs across workers, *average* utility is the same in all cities and given by:

$$(1.10) \quad V = Y \cdot \left(\sum_i L_i^{1+\frac{1}{\theta}} \cdot \frac{P_i^\beta}{Z_i}\right)^{-1}$$

In sum, conditional on the observed changes in the wage distribution, the implications for city size and aggregate output is the same as before and does not depend on θ . But the effect of

⁷ None of the substantive results here hinge on the extreme value assumption. See Kline (2010) and Busso, Gregory, and Kline (2013) for analyses with a nonparametric distribution of tastes.

local TFP, amenities, and the local housing supply elasticity on the wage distribution (and by extension on aggregate output and average welfare) depends critically on θ .

Heterogeneity in Labor Demand Elasticity: Our basic model assumes that the output elasticity with respect to labor is constant. We can relax this assumption. Specifically, suppose that total output of in a city is the sum of the output produced in different industries indexed by j :

$$Y_i = \sum_j Y_{ij}$$

where $Y_{ij} = A_i L_i^{\alpha_j} K_i^{\eta_j}$ denotes output of industry j in city i . Note that the labor and capital shares are now indexed by industry. In this case, there are two changes in the key endogenous variables. First, total employment in a city is given by:

$$L_i = \sum_j \left(\frac{\alpha_j A_{ij}}{W_i^{1-\eta_j}} \right)^{\frac{1}{1-\alpha_j-\eta_j}}$$

Second, aggregate output Y is implicitly defined by

$$1 \propto \sum_i \sum_j A_{ij}^{\frac{1}{1-\alpha_j-\eta_j}} \left(\frac{\alpha_j \bar{W}}{\bar{\alpha} Y \cdot W_i} \right)^{\frac{1-\eta_j}{1-\alpha_j-\eta_j}}$$

where $\bar{\alpha} \equiv \sum_i \sum_j \frac{Y_{ij}}{Y} \cdot \alpha_j$ is the aggregate labor share. All the other equations are the same.

Endogenous TFP and Amenities: We can also relax the assumption that TFP and amenities are exogenous. In practice, it is plausible to think that both TFP and amenities are endogenous to changes in city size. For example, a large literature in urban and regional economics posits that in the presence of agglomeration economies, A_i depends positively on L_i as denser cities are more productive. This would make our counterfactual exercise conceptually more complicated, as changes in city size would induce an endogenous feedback effect through the agglomeration economies.

In practice, our estimates of aggregate effects are not affected if the elasticity of agglomeration is constant across cities. With constant elasticity, reallocation of workers across cities has no aggregate impact, because the increases in agglomeration economies experienced by cities that grow in size are exactly offset by the losses in agglomeration economies experienced by cities that shrink in size. Empirically, the assumption of constant elasticity

appears consistent with the empirical evidence on US manufacturing (Kline and Moretti, forthcoming).

In terms of amenities, a large literature posits that amenities might depend on city size and/or density. Our baseline assumption of exogenous amenities does not require that amenities are necessarily fixed (as in the case of weather). It allows amenities --in particular public services like schools, public transit or police--to expand (contract) as the counterfactual population of the area expands (contracts), as long as the *per-capita* availability remains stable at current levels.

While this is realistic for many public services, it is possible that the per capita amount of other amenities depend on city size. This could happen, for example, if congestion is an increasing function of city size-- i.e. more people in a city mean more noise, traffic and pollution. It could also happen for the opposite reason, if more people improve urban amenities such as variety of restaurants and variety of cultural events. Glaeser (2010) has argued that cities like London, New York and San Francisco are attractive precisely because of their urban amenities stemming from high density of residents. Thus higher population density can create both negative and positive externalities.

Irrespective of the sign, the possibility of this type of endogenous amenities makes our counterfactual exercise more complicated because changes in the number of workers induce an endogenous feedback effect on residents' welfare through changes in amenities.⁸ Note that what matters is the aggregate effect. Our counterfactual exercise increases size of some cities and reduced size of other cities. If amenities decline in the first group and improve in the second group (or vice versa), the question that matters for us is the net effect in the aggregate.

To see this more clearly, consider the following extension of our model. Suppose that the production function is still given by (1.1) and welfare by (1.2) but amenities are now given by $Z_i = \bar{Z}_i L_i^{-\rho}$. Here, \bar{Z}_i denotes the component of per capita amenities exogenous to city size and $L_i^{-\rho}$ the component that varies endogenously with the size of the city. City size is given by

$$L_i \propto \left(\frac{A_i \bar{Z}_i^{1-\eta}}{P_i^{\beta(1-\eta)}} \right)^{\frac{1}{(1-\eta)(1+\rho)-\alpha}}$$
 and aggregate output and welfare by

⁸ If the elasticity of endogenous amenities with respect to city size is constant across cities, then the net effect on aggregate welfare is zero, as gains in some cities are off by losses elsewhere.

$$Y \propto \left(\sum_i A_i^{\frac{1}{(1-\eta)(1+\rho)-\alpha}} \left(\frac{\bar{Z}_i / P_i^\beta}{\sum_j (\bar{Z}_j / P_j^\beta) \cdot L_j^{1+\rho}} \right)^{\frac{1-\eta}{(1-\eta)(1+\rho)-\alpha}} \right)^{\frac{(1-\eta)(1+\rho)-\alpha}{1-\eta}}$$

$$V \propto Y \cdot \left(\sum_j (P_j^\beta / \bar{Z}_j) \cdot L_j^{1+\rho} \right)^{-1}.$$

In the end, the size and sign of the parameter ρ is an empirical question. If $\rho < 0$ then our counterfactual will imply welfare losses that will reduce the welfare benefits stemming from increased output, as it increases the size of cities that are already large, further exacerbating congestion. On the other hand, if $\rho > 0$ then our counterfactual will imply welfare gains that will magnify the welfare benefits stemming from increased output. If $\rho = 0$ then our counterfactual will be measuring welfare gains correctly.

The existing evidence indicates that the assumption that endogenous amenities are either increasing or do not depend on city size. Ahlfeldt, Redding, Sturm and Wolf (2014) and Diamond (2014) find that urban amenities slightly increase with density in Germany and the US. The most direct estimate of ρ for the US is found in Albouy (2012). He shows that quality of life in a city is positively correlated with the city population, when no controls are included. But when natural amenities such as weather and coastal location are controlled for, Albouy (2012) finds no relationship between city population and quality of life. This suggests that cities with better natural amenities are bigger (just as predicted by the equilibrium expression above for city size), but endogenous amenities are not significantly better or worse in large cities compared to small cities.⁹ If Albouy's estimates are correct, then allowing for endogenous amenities should not change our estimates of aggregate impacts very much.

Finally, it is worth highlighting an important caveat. It is in principle possible that inelastic housing supplies may contribute to the high TFP in cities like San Francisco and New York. This could happen, for example, if productivity is endogenous to college share (as in Moretti 2004 and Diamond 2013) and college workers more willing to pay high house prices. In this case, TFP would be endogenous with respect to housing supply, and our framework would not be adequate to estimate counterfactual output.

⁹ This is true within the range of city sizes observed in the data. There is of course no guarantee that if one were to significantly expand the largest cities in the US, this would remain true.

3. Data and Key Facts About the Spatial Dispersion of Wages in the U.S.

3.1 Data

The ideal data for this project would have three features: they go back in time as much as possible; they have detailed and consistently defined geocodes; and they have detailed industry definition. To approximate it, we use a combination of data sources taken from the 1964, 1965, 2008 and 2009 County Business Patterns (CBP); the 1960 and 1970 Census of Population; the 2008 and 2009 American Community Survey (ACS); and the 1964 and 2009 Current Population Survey (CPS). Since the earliest year for which we could find city-industry level data on wages and employment is 1964, we focus on changes between 1964 and 2009.

Employment, Wages and Rents: Data on employment and average wages are available at the county and county-industry level from the CBP and are aggregated to MSA and MSA-industry level. The main strength of the CBP is its fine geographical-industry detail and the fact that data are available for as far back as 1964.¹⁰ The main limitation of the CBP is that it does not provide worker level information, but only county aggregates, and it lacks information on worker characteristics. Obviously differences in worker skill across cities can be an important factor that affects average wages. In addition, union contracts may create a wedge between the marginal product of labor and the wage, as union wages may contain economic rents. We augment CBP data with MSA-level information on worker characteristics from the Census of Population, the ACS and CPS: three levels of educational attainment (high school drop-out, high school, college); race; gender; age; and union status. To purge average wage from differences in worker characteristics across cities, we calculate a *residual* wage that conditions for geographical differences in the composition of the workforce. Specifically, we use nationwide individual level regression based on the CPS in 1964 and 2009 to estimate the coefficients on worker characteristics, and use those coefficients to compute residual wages based on city averages.¹¹ We end up with a balanced sample of 220 MSA's with non-missing values in 1964 and 2009.¹² The Data Appendix provides additional information on how we defined the variables, the

¹⁰ The published tabulations of the Census of Population provide MSA level averages of worker characteristics, but the individual level data on employment and salary with geocodes is not available from the public version of the Census of Population on a systematic basis until 1980. Only a third of metro areas are identified in the 1970 Census.

¹¹ Residual wage is defined as $W - X/b$ where W is the average wage in the MSA, X is the vector of average workers characteristics in the MSA, and b is a vector of coefficients on workers characteristics from individual level regressions estimated on nationwide samples.

¹² These MSAs account for 71.6% and 72.8% of US employment in 1964 and 2009, respectively, and 74.3% and 76.3% of the US wage bill in 1964 and 2009.

limitations of the data, and presents summary statistics. Appendix Figure A1 shows that in 2009 the estimated average residual wage obtained from MSA-level data correlates well with average residual wage obtained from individual level data. (We cannot do the same for 1964, which is why we rely on MSA-level data.)

Housing Supply: Data on housing supply are from Saiz (2010). For each MSA, these data provide overall elasticity of housing supply γ_i as well as its two main determinants: land availability and land use regulations. Land use regulations are measured using the Wharton Residential Land Use Regulatory Index, originally obtained by Wharton researchers through a detailed survey of municipalities in 2007 and aggregated up at the MSA level by Saiz. It is the best available measure of differences in land use restrictions. We follow his estimates (Table 5, column 2) to divide overall supply elasticity into the part that reflects land use regulations and the part that reflects land availability.

Technology: Finally, to take the model to the data, we need to specify the technology parameters. In our baseline estimates, we assume a labor share α of 65 percent and a capital share η of 25 percent, which imply that the profit share $1 - \alpha - \eta$ is 10 percent. The assumption that the labor share is 65 percent is consistent with BEA data (BEA, 2013), data in Piketty (2014), and Karabarbounis and Neiman's (2014). The assumption that the profit share is 10 percent is consistent with Basu and Fernald's (1997) estimates of the returns to scale in U.S. manufacturing as well as with estimates in Atkeson, Khan and Ohanian (1997).

In additional estimates, we relax this assumption. First, we provide various alternative estimates under different assumptions on α and η . In these models, we either vary α and η individually or we vary the degree of returns to scale $\alpha + \eta$. Second, in separate models, we relax the assumption that technology is the same across all cities and years by allowing the technology parameters to vary by industry and over time. Because the geographical location of industries is different for different cities, this assumption allows different cities to have different technologies. In practice, we use a dataset that is analogous to the one used in the baseline analysis, but that includes separate observations (and a separate technology) for each 1-digit

industry in each city in each year. We use data on the labor share by industry in 1964 from Close and Shulenberg (1971) for 1964 and similar data for 2009 from BEA (2013).¹³

3.2 Changes in the Spatial Dispersion of Nominal Wages 1964-2009

The model in the previous section highlights the importance of wage differences across cities for aggregate output. It indicates that larger wage differences result in lower output, everything else constant. Intuitively, wage dispersion across cities reflects variation in the marginal product of labor. If labor is more productive in some areas than in others, then aggregate output may be increased by reallocating some workers from low productivity areas to high productivity ones. For example, in 2009 average nominal wages in San Jose, CA were twice as large as nominal wages in Brownsville, TX, presumably because the marginal product of labor in San Jose is twice as large. If some workers were moved from Brownsville to San Jose, aggregate GDP would increase because more workers would have access to whatever productive factor generates high productivity in San Jose. In principle, aggregate output is maximized when the marginal product of labor is equalized across locations.

Empirically, the spatial distribution of nominal wages across US metropolitan areas is significantly more dispersed in 2009 than it was in 1964, suggesting a negative effect on output growth. Figure 1a plots the weighted distribution of the unconditional average wage in a MSA in 1964 and 2009 (after removing the mean US wage in each year), where the weights are MSA employment in the relevant year. It is clear that the 2009 distribution is more significantly dispersed. It is also clear that the right tail -- which includes cities with average wages that are 50% above the mean -- has become thicker.¹⁴ The bump of the right tail includes New York, San Francisco and San Jose.

Table 1a quantifies the change in the dispersion in average nominal wage. Panel A indicates that the employment-weighted standard deviation (column 1), interquartile range (column 2), and the range (column 3) of the log average MSA wage increased significantly from 1964 to 2009 (by .07 log points, .10 log points, and .38 log points respectively). Panel B controls for the average wage in nine Census divisions and it suggests that increases in wage dispersion is not just a regional phenomenon, but it occurs even within Census divisions. Indeed, controlling

¹³ No historical data exist on capital share by industry or by city. In both years, we retain the assumption of a 10% profit share (Basu and Fernald, 1997).

¹⁴ Weighting is empirically important. The unweighted distribution shows a more limited increase in dispersion (see Appendix Figure A2). As our model makes clear, the weighted distribution is the relevant one for our purposes.

for regional wage differences generates a *larger* increase in the wage dispersion, so regional wage differences have declined over time.

Panel C shows that a non-trivial part of the increase in dispersion is due to three large cities with high output growth over the last 50 years: New York, San Francisco, and San Jose. Dropping these three cities has a significant effect on the right tail of the distribution. The standard deviation and the range of the log average wage when we exclude these three cities increase by much less from 1964 to 2009.

These findings are not driven by differences in observable worker characteristics across cities. Figure 1b and Table 1b present the spatial dispersion of the average of the *residual* wage in each MSA.¹⁵ Controlling for changes in worker composition does not alter the picture of increased spatial dispersion between 1964 and 2009. The picture that emerges indicate that (i) spatial dispersion has increased significantly; (ii) such increase is not all concentrated in one specific region; and (iii) New York, San Francisco and San Jose account for an important part of such increase.

These findings are generally robust. First, all the results are identical if we use 2007 data (pre-recession) instead of 2009.¹⁶ Second, our approach of controlling for workers characteristics assumes that the effect of workers characteristics is the same everywhere in the country, but it is possible that the return to these characteristics (such as education) varies across cities (Dahl, 2003). To see whether this matters empirically, we estimate models where we allow the effect of workers characteristics on wages to vary by region or by state. When we do this, the resulting spatial distribution of wage residuals is very similar to that shown in Table 1b.

A potential concern is that we cannot control for unobserved differences in worker ability. It is possible that average unobserved ability differs between cities, and that some of the documented wage differences across cities are not differences in the marginal product of labor, but difference in the *quality* of labor.

We cannot completely rule out the possibility of unobserved worker heterogeneity. However, three considerations are worth mentioning. First, the fact that the unconditional

¹⁵ The five cities for which the difference between unconditional and conditional wages is the largest are, Madison WI; Ann Arbor, MI; Boston, MA; Champaign-Urbana, IL; State College, PA. In other words, controlling for education and other workers characteristics has the largest impact in university towns and other cities with very high density of college educated workers. By contrast, the five cities for which the difference between unconditional and conditional wages is the smallest are McAllen, TX; Brownsville, TX; Visalia, CA; Yakima, WA; and Bakersfield, CA. Controlling for education and other workers characteristics has the smallest impact in cities that have a labor force with low levels of schooling and high levels of minority workers.

¹⁶ As explained in the Appendix, to increase the sample size, our 2009 data actually includes 2008 and 2009.

distribution (Figure 1a) is basically the same as the distribution conditional on observable worker characteristics (Figure 1b) should alleviate the concern at least in part.

Second, recent evidence based on longitudinal data that follow workers moving from low wage cities to high wage cities indicates that this problem may be limited once education is controlled for. Baum-Snow and Pavan (2012), for example, find that sorting on unobserved ability within education group contributes little to observed differences in wages across cities of different size. Similarly, De La Roca and Puga (2012) find that workers in cities that are bigger and have higher wages do not have higher unobserved initial ability, as reflected in individual fixed-effects. These findings are consistent with Glaeser and Mare (2001), who show that workers who move from low wage areas to high wage areas experience significant wage increases and that this is not just the result of sorting by ability. We also point out that what matters for our analysis is not merely the possibility of differences in unobserved ability in a cross-section of cities. Rather is whether these differences have changed differentially over time.

Third, we have explored the relationship between worker ability and nominal wages. Specifically, we have used NLSY data to relate the average AFQT scores to the nominal wage across metropolitan areas. This data indicates that workers in high nominal wage MSAs tend to have higher AFQT scores, but the correlation attenuates and becomes statistically insignificant once we introduce controls for education, race, and ethnicity.¹⁷

The flipside of the increase in the dispersion in wages is an increase in the dispersion in housing costs, since in equilibrium workers need to be compensated for housing costs. Panel A in Appendix Table A2 shows that the dispersion in average rent has increased between 1964 and 2009. Rents are a good approximation to the user cost of housing. In panel B we show the corresponding figures for housing prices. The increase in the spatial dispersion of housing prices is larger than that of housing rents.

4. Empirical Findings

We now take the model to the data. First, we decompose aggregate GDP growth into the contribution of each US city and compare it with a naïve “accounting” calculation (section 4.1). Second, we turn to the increased dispersion of wages and calculate how much larger US GDP

¹⁷ We first regress log AFQT scores on log nominal wages. We then replicated the same regression controlling for the same vector of controls used in panel B of table 1b. Both regressions are weighted by MSA employment. While the coefficient is positive in the first regression, it is statistically indistinguishable from zero in the second regression. However, the small sample size precludes definitive conclusions.

would be in the counterfactual where the spatial dispersion of wages is fixed from 1964 to 2009 (section 4.2). We then turn to the causes of increased dispersion of wages to discuss its welfare implications (section 4.3). Finally, we discuss limitations of our approach (section 4.4).

4.1 Local Growth and Aggregate Growth

Equation (1.7) allows us to calculate the contribution of each city to aggregate growth in 1964 and 2009. This calculation is presented in Table 2 and Figures 2a-2e. Figure 2a plots the percentage contribution of the 220 cities to aggregate growth from 1964 to 2009 (on the y-axis) against the growth of local GDP as a percentage of aggregate GDP growth over the same period (on the x-axis). To be clear, the calculation on the y-axis is based on the model (specifically on equation (1.7)) while the x-axis is the growth in local GDP as a ratio of aggregate GDP growth.¹⁸ We call the latter the naïve “accounting” calculation. The solid line is the 45 degree line so cities that lie above the 45 degree line contribute more to growth than is apparent from their measured GDP growth, and cities below the 45 degree line contribute less to growth than suggested by their output growth. If all the observations lie on the 45 degree line, the growth rate of aggregate GDP would simply be given by the weighted average of local GDP growth.

The first feature that is apparent in Figure 2a is that the dispersion of the accounting measure of the contribution of each city is much wider than the actual contribution. The range of the accounting calculation of the contribution of a city to aggregate growth is 20 percent while the range of the model based calculation is only 5 percent.

The second and most important feature of Figure 2a is that there are sizable and systematic differences between local growth and local contribution to aggregate growth. For example, growth of New York’s GDP was 12 percent of aggregate output growth from 1964 to 2009. However, viewed from the lenses of the Rosen-Roback model, New York was only responsible for less than 5 percent of aggregate output growth. The difference is because much of the output growth in New York was manifested as higher nominal wages, which increased the overall spatial misallocation of labor. On the other extreme, Detroit’s GDP fell dramatically from 1964 to 2009. Although one might expect the contribution of Detroit to be negative because of the

¹⁸ The “accounting” calculation is based on the accounting identity $\frac{y_{t+1}}{y_t} = \sum_i \frac{y_{i,t+1}}{y_{i,t}} \cdot \frac{y_{i,t}}{y_t} \cdot L_{i,t}$ where the t and $t+1$ subscripts denote time, y denotes aggregate GDP per worker (in the country), y_i is GDP per worker in city i , and L_i the employment share in city i . The “contribution” of city i is measured by $\frac{y_{i,t+1}}{y_{i,t}} \cdot \frac{y_{i,t}}{y_t} \cdot L_{i,t}$.

decline measured local output, the net contribution of Detroit to aggregate output growth is positive. The difference in the case of Detroit is because much of the decline in local GDP in Detroit was driven by a decline in nominal wages. And in Detroit, nominal wages in 1964 were significantly higher than the nationwide mean so the decline in the nominal wage from 1964 to 2009 lowered the overall wage dispersion which increases aggregate output. Other cities display large differences: Chicago and Los Angeles, for example, are well above the 45 degree lines. While their contribution to aggregate growth as calculated from equation (1.7) is not unlike New York's contribution, the growth of their GDP as a fraction of overall growth is much smaller.

Overall, Figure 2a shows that the relation between local growth and local contribution to aggregate growth is positive, but with an elasticity that is much less than one. A regression of the variable on the y-axis on the variable on the x-axis yields a coefficient (standard error) of .295 (.018), with an intercept equal to .320 (.033). The slope is statistically different from one and the intercept is statistically different from zero. Cities with large positive shocks to their local economy tend to contribute less to aggregate growth than their local gains would suggest. At the same time, cities with large negative shocks tend to contribute more than their local losses would suggest. This discrepancy between local growth and local contribution reflects changes in each city relative wage.

Figures 2b-2e and Table 2 separately present the contribution of four groups of cities. Figure 2b presents the actual vs. the accounting calculation of New York, San Francisco, and San Jose to aggregate output growth from 1964 to 2009. All three cities lie significantly below the 45 degree line. Although local output grew rapidly in all three cities, so did the gap between local wages and the nationwide wage. The first row in Table 2 indicates that although local GDP growth was almost 20 percent of aggregate US output growth, the actual contribution of these three cities was much lower, at 6 percent of US output growth.

Figure 2c shows the contribution of 37 cities in the Rust Belt. As can be seen, all the Rust Belt cities lie above the 45 degree line: the actual contribution of Rust Belt cities is larger than suggested by observed changes in local GDP. What is driving this discrepancy is that nominal wages in Rust Belt cities were typically above the nationwide mean in 1964. And since 1964 wages have fallen and have thus narrowed the gap between wages in the Rust belt and the nationwide mean. What is perhaps more surprising is that although local GDP growth is negative in every Rust Belt city, the actual contribution of every Rust Belt city to aggregate growth is positive. Although the decline in labor demand caused by the decline of manufacturing presumably implies that the contribution of the Rust Belt cities to aggregate

growth would be negative, the allocative effects of the sharp decline in the wage gap has a larger effect on aggregate growth. Table 2 shows that the Rust Belt cities contributed as much as New York, San Jose, and San Francisco (taken together) to aggregate output, despite the sharp decline in GDP in the Rust Belt cities.

Figure 2d presents the contribution of 86 Southern cities. In the period under consideration, the South of the US has grown more rapidly than the rest of the country. Washington, DC, Houston, Atlanta, and Dallas are among five fastest growing cities in the US (the fastest growing city is New York). All cities lie significantly below the 45% line because the gap in local wages and the nationwide wage increased in all these cities. Therefore, the contribution of the large Southern cities to aggregate growth is less than suggested by their output growth. The fact that relative wages increased in the large southern cities also suggests that the standard narrative that growth in these cities was driven by improved amenities (hot weather became more tolerable with air conditioning) and cheap housing is not the entire story. If the only change in the South was that amenities have improved or housing became cheaper, then relative wages should have fallen in these cities whereas the opposite is true. Taken together, Southern cities were responsible for 42 percent of aggregate growth in the US (Table 2). This is sizeable to be sure, but 20 percentage points lower than what one might infer from the observed growth of GDP in the Southern cities.

Figure 2e presents the contribution of the remaining large US cities. This group includes 19 large cities with 2009 employment above 600,000 that are not in any of the previous three groups. Here, the story is more mixed. There are cities where the observed local growth almost exactly measures the actual contribution. These are cities such as Boston, Portland, and Salt Lake City. There are also cities where the growth contribution is larger than suggested by local growth. These are cities such as Chicago, Los Angeles, and Philadelphia where relative wages have fallen and the gap in the marginal product of labor relative to the rest of the country has narrowed. Finally, there are also cities where the growth contribution is smaller than suggested by the local output growth numbers. For example, Phoenix, is one of the fastest growing metro areas in the country; based on the accounting measure, GDP growth in Phoenix “accounts” for six percent of aggregate US growth. Yet, much of this growth has accompanied by a *decline* in wages in Phoenix, which in the framework of the Rosen-Roback model must be driven by a decline in relative housing prices or an improvement in relative amenities. And since wages in Phoenix were already below the nationwide mean in 1964, the further decline in wages increases the wage gap. Las Vegas and Riverside have similar experiences. Essentially, Phoenix, Las

Vegas and Riverside have attracted many residents because of good weather and abundant supply of cheap housing but this reallocation results in a loss in aggregate output because it has brought more people working in cities where the marginal product of labor is low. This effect is reminiscent of the Dutch disease in two-sector models of growth.

The bottom line is that almost three quarters of aggregate US output growth from 1964 to 2009 was driven by local forces in southern US cities and the group of "large" 19 cities. And despite the large difference in local GDP growth between New York, San Jose, and San Francisco and the Rust Belt cities, both groups of cities had roughly the same contribution to aggregate output growth (about 6 percent).

In Table 3 we probe the robustness of our estimates using different assumptions on technology. Recall that our baseline estimates assume that $\alpha = .65$ and $\eta = .25$ in all cities in both years (column 1). In columns 2 and 3, we keep the returns to scale constant and alter α or η . The estimates are almost identical to the baseline estimates. In columns 4 to 7 we alter the labor or capital share to vary the returns to scale. In columns 4 and 5, we increase return to scale, as $\alpha + \eta$ increases from .9 to .95. In columns 6 and 7 we alter the labor or capital share to decrease the returns to scale -- $\alpha + \eta$ decreases from .9 to .85. Entries are virtually unchanged.

So far we have constrained the technology to be the same in all cities and industries. Next, we relax our assumptions on technology by allowing technology to vary across cities and years. Specifically, we allow labor and capital shares in 1964 and 2009 to be different in different industries. Because the geographical locations of industries are not the same, this allows different cities to have different technologies. In practice, we use a dataset that is analogous to the one used in the baseline analysis, but that includes separate observations for each 1-digit industry in each city in each year. We assume that workers can move freely across industries within each city, so that the wage is the same. The entries in column 8 indicate that the results are not very sensitive to this generalization.

We have performed several additional checks, and found our results to be generally robust. For example, in some models residual wage is estimated using models where the coefficient of workers characteristics is allowed to vary not just by year, but also by state. Results did not change significantly. We have also re-estimated our models dropping the two cities that in Figure A1 are outliers, and found similar estimates.¹⁹

¹⁹ We have also re-estimated our 2009 model dropping the restaurant sector, as one where minimum wage workers are particularly prevalent and therefore the assumption that equates wages with marginal product of labor may be violated. The correlation of the share that each city contributes to 2009 output with and without the restaurant sector is .99. We can't do the same for 1964, since industry definition in 1964 is less disaggregated.

4.2 Wage Dispersion and Aggregate Growth

Equation (1.7) decomposes the growth rate of aggregate output into two components: growth of local TFP and change in the spatial dispersion of wages. It indicates that increases in the spatial dispersion of wages negatively affect aggregate growth: for a given local TFP growth, a more dispersed spatial wage distribution results in slower growth. Empirically, we have seen that the spatial dispersion of wages across US cities increased significantly from 1964 to 2009 --- the standard deviation, for example, is now double relative what it used to be in 1964.

We now quantify the effect of this increase in wage dispersion on the rate of growth of aggregate output between 1964 and 2009 and on the level of output in 2009. We estimate counterfactual output under the scenario where the dispersion of wages across cities remained constant between 1964 and 2009. Specifically, we calculate the counterfactual where the relative wage of a city in 2009 is equal to the relative wage of the same city in 1964. We take local TFP in each city as fixed and allow labor and capital to endogenously reallocate across cities in response to the change in the distribution of local housing supply and amenities. Clearly wages are an endogenous variable. As we have seen, they are determined by local TFP, amenities and elasticity of housing supply (equation (1.4)). But the effect of changes in the wage dispersion on aggregate output growth does not depend on the sources of wage dispersion. (The effect on welfare does depend on the source of wage dispersion. We take up the question of the exact mechanism underlying the change in the spatial wage dispersion in the next section.)

In terms of output growth, when we take equation (1.7) to the data, we find that the growth of local TFP boosts aggregate GDP by 2.5 percent a year from 1964 to 2009, holding the spatial dispersion of wages fixed. The increased spatial dispersion of wages lowers aggregate GDP growth by 0.3 percent a year, holding constant local TFP. The net effect of these two forces is that aggregate GDP grew by 2.2 percent a year from 1964 to 2009. In other words, under the counterfactual scenario where wage dispersion did not increase in the U.S., aggregate yearly GDP growth from 1964 to 2009 would have been 0.3 percentage points higher.

In terms of output level, the increase in the spatial dispersion of wages resulted in a significantly lower level of output in 2009. This effect is quantified in Table 4. The first row indicates that if the spatial dispersion of relative wages had not changed, 2009 U.S. GDP would be 13.5% higher. Given that US GDP in 2009 was 14.5 trillion, this implies an additional annual aggregate income of \$1.95 trillion. Given a labor share of .65, this amounts to an increase of

\$1.27 trillion in the wage bill, or \$8775 additional salary per worker (if number of workers was fixed).²⁰ More than half of US workers would move under this scenario (column 2).

In the second row of Table 4, we set the distribution of nominal wages in 2009 equal to its 1964 level only in New York, San Francisco and San Jose. Remember that the increase in relative wages from 1964 to 2009 was particularly pronounced in these cities. In addition, these cities are among the largest cities in the US in terms of TFP so the effect on aggregate output growth of the change in the change in the wage in these three cities is largely to be large. Aggregate output would increase by 13.2% if the relative average wage in only these three cities is set to their 1964 level. 54% of U.S. workers would relocate.²¹

The third row illustrates the effect on aggregate output when the distance from the mean wage in the Rust Belt cities is set to gap in 1964. As can be seen, the effect is small, as aggregate 2009 output increases by 0.5% and only 9% of workers relocate. The last row shows the effect on aggregate output when the distance from the mean wage in Southern cities is set to gap in 1964. Row 3 shows that if the distance from the average wage in Southern cities were set at the 1964 gap, aggregate 2009 output would *fall* by 0.4%.

The changes in the economic geography of the US implied by Table 4 are massive and probably not realistic. Changing the geographical location of American workers to the point that brings wages back to their 1964 level would likely take several decades. One way to see how extreme implied by this scenario is to compare the implied mobility rate with the one observed in reality. Consider that less than 20% of workers change MSA every 10 years. By comparison, the scenario in row 1 of Table 4 involves the relocation of more than half of the US work force.

Table 5 shows the equivalent of Table 4, but for partial adjustment. We scale partial adjustment based on the fraction of movers. For example, the second row in the table shows that if 2009 wages were set so that only 50% of workers were to relocate, the output gain in 2009 is 13.2%. The other rows show that if 2009 wages were set so that only 40%, 30%, 20% or 10% of workers were to relocate, the output gain would be respectively 11.8%, 9.4%, 6.5%, and 3.4%.

We consider the scenario where 20% of workers change MSA -- corresponding to the counterfactual shown in the fifth row of Table 5--as our benchmark scenario, as it is the closest to the typical mobility rate that we observe over a decade.

²⁰ The salary increase would be smaller if more workers decide to enter the labor market in response to the higher salary.

²¹ The size of these three cities would grow. It is important to understand, however, that in general equilibrium the spatial relocation of labor would affect not only these three cities, but all cities in the U.S.

Table 6 shows counterfactual employment for selected cities under full adjustment and partial adjustment. In particular, in column 1, counterfactual employment is computed setting 2009 relative wage to 1964 levels in all cities (first row of Table 4). In column 2, counterfactual employment is computed moving 2009 relative wage toward their 1964 levels in all cities up to the point where 20% of U.S. workers change MSA (row 5 of Table 5).

By a vast margin, New York is the city that would experience the largest percentage increase in employment: a staggering 787% increase in the case of full adjustment. San Jose and San Francisco would grow by more than 500%, while Austin would increase by 237%. All these cities are important innovation clusters and have experienced rapid wage growth since 1964 mostly driven by human capital intensive industries. Surprisingly, Fayetteville is also in the top group. What distinguishes this MSA is the fact that its economy has changed enormously over the past 3 decades due to the location of Walmart headquarters. The median city, Sheboygan, WI would lose 80% of its employment. The bottom of the table reports the cities that would experience the largest decline in employment. This group includes Rust Belt former manufacturing centers, like Mansfield OH, Muncie, IN and Flint, MI. Under our counterfactual scenario, virtually all of Flint's workers would move and relocate to other cities.

Column 2 shows the counterfactual employment for selected cities under the more plausible intermediate scenario where 20% of workers change city of residence. New York remains the city that would experience the largest percentage increase in employment, but the increase is only 179%. San Jose, San Francisco, Fayetteville and Austin would grow by 149%, 147%, 118% and 102%, respectively. The median city Sheboygan, would lose a third of its employment. The bottom of the table indicates that 78% of Flint's workers would move and relocate to other cities.

Three considerations are worth keeping in mind. First, these are intended to be long term benchmarks. They are based on the assumption that as the population expands in an area, local services also expand to keep the per-capita availability of schools, parks, public transit and other public amenities stable at their current levels. Thus, one should not think of these counterfactuals as taking place overnight and holding fixed public services. Rather, one should think of these counterfactuals taking place slowly over the long run, matched with a steady increase in the supply of public services so that the *per-capita* level of public services is unchanged.

Second, while the counterfactual employment for the top group of cities in column 2 imply city sizes that are very large, they are not completely implausible. For example, the Association of Bay Area Governments (which is made of all municipalities in the San Francisco

Bay Area) has recently adopted a formal economic development plan for the region that calls for the addition of enough housing units to increase the region's population by 80% in 2030 (ABAG, 2013). This increase is smaller than the one estimated in column 2 of Table 6 for the San Francisco MSA, but not too far off.

Third, these estimates are obtained assuming that the total number of workers in the US is fixed. In reality, if wages were to rise of average, total employment is likely to increase due to international migration and increased domestic labor supply. This would further increase counterfactual output. Thus, our estimates of output gains are to be interpreted as a lower bound.

In Appendix Table A3 we probe the robustness of our estimates using different assumptions to calibrate the model parameters. In rows 2 and 3, we keep the returns to scale constant and alter α or η . In rows 4 to 7 we alter the labor or capital share to vary the returns to scale. We find that the results are not sensitive to changes in labor or capital share for a given degree of return to scale. But they are quantitatively sensitive to the degree of decreasing return to scale. The closer the sum $\alpha + \eta$ is to 1, the larger the output gain. This makes intuitive sense, because the sum $\alpha + \eta$ governs the returns to scale. With $\alpha + \eta$ close to 1 our technology approaches constant returns to scale and there is the most productive cities attracting an increasingly larger share of the economic activity of the country. Finally, in the bottom row, we allow labor and capital shares in 1964 and 2009 to be different in different industries and years. Since cities have different shares of each industry, this models allows technology to vary across cities and years, as a function of their industry mix.

4.3 Sources of Wage Dispersion: Housing Supply vs. Amenities

We have shown that the spatial dispersion of nominal wages has increased significantly over the past 50 years and, as a consequence, aggregate growth and aggregate output are lower than what they could have been. However, we have been silent on what has caused the increase in wage dispersion and on the implications for welfare. Formally, we have shown that the difference between welfare and output is simply the weighted average of the ratio of housing prices to local amenities. Understanding how changes in housing prices and amenities have affected wages is thus crucial to understand the implications of changes in wages for welfare.

In other words, we need to determine *why* U.S. labor is not flowing to high wage cities to a larger degree. Our calculations of the counterfactual output in the previous section did not depend on the specific reason for the increased spatial dispersion in wages. But to understand the implications for welfare, we need to understand what has been increasingly constraining labor

supply to high wage cities in the U.S. In our setting, labor supply to a city depends on two exogenous factors ---amenities and elasticity of housing supply --- with opposite implications for worker welfare.

Intuitively, if labor is not moving to high wage cities like San Francisco or New York because of undesirable amenities – for example, workers may find these cities crowded, noisy and polluted -- then increasing their size will increase aggregate output but not aggregate welfare. On the other hand, if labor is not moving to cities like San Francisco or New York due to housing supply constraints caused by land use regulations, then increasing their size will increase aggregate output and aggregate welfare.

This possibility is consistent with anecdotal evidence on the evolution of land use regulations over the past half century. Glaeser (2014), among others, points out that since the 1960's, expensive coastal U.S. cities have gone through a property rights revolution which has significantly reduced the elasticity of housing supply: “In the 1960s, developers found it easy to do business in much of the country [...]. In the past 25 years, construction has come to face enormous challenges from any local opposition. In some areas it feels as if every neighbor has veto rights over every project.”²²

We now examine which of these two factors –amenities or housing supply restrictions created by land use regulations---have contributed the most to the output losses uncovered above.

(A) Amenities: The effect of the distribution of amenities on aggregate output depends on whether amenities have improved more in high wage cities or in low wage cities. If amenities have improved by more in high wage cities, this lowers the dispersion of the nominal wage across cities and, *ceteris paribus*, increases aggregate output.

Consistent urban economics literature, we use the spatial equilibrium condition (equation (1.2)) to measure amenities: $Z_i \propto W_i/P_i^\beta$. This condition indicates that local amenities are proportional to the difference between properly weighted housing rents and nominal wages, where the weight on housing rents β reflects the share of housing in total expenditures. We set the housing share β equal to 0.32 from Albouy's (2012) estimates.²³ Albouy (2012) shows that

²² Glaeser also points to political economy causes of this trend: “To most residents, a new project is nothing but a bother. They don't care about the welfare received by the new resident, or the benefits earned by the builders or by the employers who have to pay lower wages when housing costs are lower. Moreover, unaffordable housing isn't a problem to most homeowners — it represents an increase in the value of their biggest asset.” (Glaeser, 2014)

²³ Following Albouy (2012) we multiply wages by 0.52 to account for taxes and transfers. Note that amenity levels are not identified because we do not know the absolute value of welfare.

this measure of local amenities is highly correlated with available measures of specific amenities (such as weather and crime) and with existing indices of the quality of life.

Table 7 quantifies the role played by changes in amenities.²⁴ In the top row, we compute counterfactual output under the assumption that the level of amenities in 2009 is set equal to its 1964 level. To obtain this counterfactual, we proceed in two steps. We first use equation (1.4) and compute what wages would be in 2009 had amenities in each city stayed at 1964 levels (holding TFP and housing supply constant). We then allow workers and capital to reallocate and compute counterfactual employment and output.

The results in the first row of Table 7 show that counterfactual output is higher than observed output, but only marginally. If the level of amenities in 2009 was equal to its 1964 level, 2009 output would grow by only 1.6% and less than 10% of workers would move.

In rows 2 to 4, we repeat the same exercise changing amenities levels only in selected cities. Row 2 shows that changes in amenities in New York, San Francisco and San Jose between 1964 and 2009 had a positive impact on aggregate output, but the effect is quantitatively small.²⁵

Row 3 performs the same exercise for the Rust Belt. Our estimates indicate that, unsurprisingly, amenities in Rust Belt cities worsened from 1964 to 2009. Changing amenities back to their 1964 level would further lower wages in the Rust Belt and slightly increase the overall wage dispersion. Row 3 shows that aggregate output would fall under this scenario, although the magnitude is trivial.

In row 4 we look at the South. Empirically, amenities have improved in Southern cities from 1964 to 2009. This is plausible, and likely reflects air conditioning, and the general improvement in quality of life in the South. Rolling amenities back to their 1964 level would increase wages in the South and slightly reduce the overall wage dispersion. Aggregate output would increase under this scenario, although the estimate in row 4 indicates that the effect is very small. Here the improvement in amenities experienced by Southern cities increases aggregate welfare, but this effect is slightly offset by the decline in aggregate output.

²⁴Appendix Table A4 shows that the spatial dispersion of amenities has increased between 1964 and 2009, although the increase in the spatial dispersion is significantly less than that observed for wages

²⁵ While crime, cultural amenities and quality of life in general are generally thought to be better to have improved in New York, San Francisco and San Jose since the 1990s, the evidence in row 2 suggests that the post 1990s improvement in amenities have offset the decline in amenities prior to the 1990s. So here, the change in amenities in New York, San Francisco and San Jose has two effects on welfare. First, it directly lowers the average level of amenities. Second, it increases the nominal wage in these cities, increases the overall wage dispersion, and lowers aggregate output.

In sum, we conclude that amenities have changed differentially across US cities. But the overall effect across all cities of changes in the distribution of amenities is limited and cannot explain but a small fraction of our counterfactual output gains.

(B) Housing Supply: In our model, the equilibrium housing price is given by $P_i \propto (A_i Z_i^{1-\eta})^{\frac{\gamma_i}{(1-\eta)(1+\beta\gamma_i)-\alpha}}$. This says that higher housing prices can be driven by higher local TFP, better amenities, and more inelastic housing supply (higher γ_i).

Based on Saiz's estimates, New York, San Francisco and San Jose have some of the most inelastic housing supplies in the country (high γ_i). Specifically, San Francisco is at the 99th percentile of the inverse elasticity distribution, while New York and San Jose are at the 96th percentile. Saiz shows that this is due to a combination of geographical features and restrictions to housing supply due to land use regulations, as measured by the Wharton Residential Land Use Regulatory Index.

We cannot measure land use restrictions in 1964, because the Wharton survey does not go back in time. Instead, in Table 8, we estimate counterfactual output under the assumption that land use regulations in New York, San Francisco and San Jose are set equal to the level of regulations in the median US city. Thus, our counterfactual takes as given geographical factors that can affect housing supply, and only changes factors that are set by policy.

To obtain this counterfactual, we proceed in three steps. First, we use Saiz (2010) coefficients (Table 5 column 2 in his paper) to estimate the elasticity of housing supply in New York, San Francisco and San Jose if land use regulations in these three cities were equal to the level of regulations in the median US city, holding constant geography. The resulting counterfactual elasticity of housing supply is mechanically higher in these three cities. Second, we use this counterfactual elasticity to estimate the counterfactual levels of housing prices and wages in New York, San Francisco and San Jose holding local TFP and amenities constant at 2009 levels, along with the counterfactual employment levels. Empirically, we find that counterfactual wages are on average 25% lower in the three cities and employment is higher. This is not surprising: because counterfactual housing supply is more accommodating, in equilibrium more workers can move to these three cities from the rest of the US. Empirically, San Francisco is the city that grows the most in this counterfactual, followed by New York and San Jose. Third, we compute the counterfactual output that is generated by this new allocation of labor.

The first row in Table 8 indicates that this would significantly speed up growth. The difference between the actual and counterfactual annualized output growth rate between 1964 and 2009 is .21%. This would induce 30% of workers to relocate, and would increase 2009 output level by 9.7%. Comparing this figure with the corresponding estimate in Table 4 (13.5%), we conclude that this change in supply elasticities accounts for more than two thirds of the overall output gains.

The second row of the table focuses on the role played by land use regulations in the South. Housing supply is generally rather elastic in Southern cities. This reflects abundant land and permissive land use regulations. We estimate counterfactual output under the assumption that land use regulations in the South are set to the level of New York, San Francisco and San Jose, holding constant land availability in the South. More stringent regulations would result in higher wages and lower employment in the South. The entry shows that in turn, US output would be 3% lower in this counterfactual scenario.

We note that our estimates are sensitive to the assumption of perfect mobility. In the theory section, we have shown how preferences for location may reduce the effect of changes in amenities or housing supply, although they do not alter the estimates of the overall effect of changes in relative wages. The key parameter in this case is the dispersion parameter, which governs the strength of preference for location. Stronger preferences for location induce some individuals to optimally choose cities where real wages net of amenities are low. To our knowledge, there are only two empirical estimates of this parameter based on MSA-level data, although neither fits our setting perfectly. Serrato and Zidar (2014, Table 5) estimate this parameter to be in the range .47 - .75, while Diamond (2013, Table 3) estimates the parameter to be .57 for college graduates and .27 for workers with lower education. If we use the largest value of the parameter in Serrato and Zidar's -- .75 -- we find output gains that are significantly smaller. For example, the estimate in row 1 of Table 8 drops to 1.6%. In this case, employment in New York, San Francisco and San Jose increase only by 54%, 50%, and 31% respectively. We note however, that both Serrato and Zidar's and Diamond's parameters are likely to be conservative for our setting, as they are obtained using 10 year changes or less. A longer time horizon would likely imply more mobility and yield larger estimates.

4.4 Caveats and Limitations.

This paper highlights the possibility of output and welfare losses stemming from an inefficient geographical allocation of labor. The number we present should not be taken as

precise estimates of the losses but rather as guidance on the general order of magnitude of the losses, as they are based on a number of untestable assumptions.

First, our findings depend on specific assumptions on technology. While our estimates are qualitatively robust to alternative technology parameters, we have shown that they are quantitatively sensitive to the assumed degree of returns to scale (Appendix Table A3).

Second, we use residual wages as a measure of the marginal product of labor. This requires that differences across cities in unobserved worker characteristics have not changed over time, or, if they have changed, they have changed in ways that are uncorrelated with nominal wages. While this might not be true, there is little we can do to relax this assumption, as detailed data on worker cognitive ability are not available at a scale large enough to allow for a city-level analysis. Failure of this assumption may lead us to overestimate potential benefits of geographical reallocation of labor. In particular, if workers in MSA's with high nominal wages have higher IQ than workers in MSA's with low nominal wages after conditioning on education and other characteristics, then the documented spatial dispersion in nominal wages overestimates the true degree of dispersion. If, in addition, the amount of unobserved ability has increased more in MSA's with high nominal wages than in MSA's with low nominal wages, then the estimated counterfactual output gains reported in the paper are too large.

Third, we have made restrictive assumptions on the relationship between TFP and city size; and the relationship between amenities and city size. A large literature in urban economics indicates that TFP might not be exogenous, but could depend on the size or the density of a city. Similarly, it has long been posited that local amenities can depend on city size and/or density. Our assumptions don't rule out these possibilities, but restrict the relationship between TFP and employment and the relationship between amenities and employment. Recall that we have assumed that the elasticity of agglomeration and the elasticity of amenities is constant across cities. With constant elasticity, reallocation of workers across cities has no aggregate impact on aggregate productivity or aggregate amenities, because the changes experienced by cities that grow in size are exactly offset by changes experienced by cities that shrink in size. As noted above, the assumption of constant elasticity for TFP is consistent with Kline and Moretti, forthcoming; the assumption of constant elasticity for amenities is consistent with Albouy (2012). However, we stress that the estimates in both Kline and Moretti (forthcoming) and Albouy (2012) are based on ranges of city size historically observed in the U.S. data. There is no guarantee that the same estimates extend to city sizes that are significantly larger than the ones observed in the data.

Fourth, we have assumed that workers can freely move across industries. This assumption is useful because cities have distinct industry specialization. Thus, spatial reallocation of labor also implies industry reallocation. For example, scaling up employment in New York, San Francisco and San Jose implicitly requires increasing the number of workers in finance and high tech, since tradable sector employment in these three cities is heavily concentrated in finance and high tech. The assumption of inter-industry mobility is clearly false in the short run. For example, it would be hard to relocate a Detroit car manufacturing worker to a San Francisco high tech firm overnight. On the other hand, the assumption is more plausible in the long run, as workers skills –especially the skills of new workers entering the labor market --- can adjust. In this respect, it is important to note that not all the workers need to adjust, because not all the workers are spatially reallocated in our counterfactual exercises. In addition, not all workers are employed in the tradable sector. While wages are set in the tradable sector, two third of the labor force is employed in the non-tradable sector, which is arguably much less specialized.

5. Policy Implications

We find that three quarters of aggregate U.S. growth between 1964 and 2009 was due to growth in Southern US cities and a group of 19 other cities. Although labor productivity and labor demand grew most rapidly in New York, San Francisco, and San Jose thanks to a concentration of human capital intensive industries like high tech and finance, growth in these three cities had limited benefits for the U.S. as a whole. The reason is that the main effect of the fast productivity growth in New York, San Francisco, and San Jose was an increase in local housing prices and local wages, not in employment. In the presence of strong labor demand, tight housing supply constraints effectively limited employment growth in these cities. In contrast, the housing supply was relatively elastic in Southern cities. Therefore, TFP growth in these cities had a modest effect on housing prices and wages and a large effect on local employment.

Constraints to housing supply reflect both land availability and deliberate land use regulations. We estimate that holding constant land availability, but lowering regulatory constraints in New York, San Francisco, and San Jose cities to the level of the median city would expand their work force and increase U.S. GDP by 9.5%. Our results thus suggest that local land use regulations that restrict housing supply in dynamic labor markets have important externalities on the rest of the country. Incumbent homeowners in high wage cities have a private incentive to

restrict housing supply. By doing so, these voters de facto limit the number of US workers who have access to the most productive of American cities.

For example, Silicon Valley---the area between San Francisco and San Jose---has some of the most productive labor in the globe. But, as Glaeser (2014) puts it, “by global urban standards, the area is remarkably low density” due to land use restrictions. In a region with some of the most expensive real estate in the world, surface parking lots, 1-story buildings and underutilized pieces of land are still remarkably common due to land use restrictions. While the region’s natural amenities---its hills, beaches and parks---are part of the attractiveness of the area, there is enough underutilized land within its urban core that housing units could be greatly expanded without any reduction in natural amenities. Our findings indicate that in general equilibrium, this would raise income and welfare of *all* US workers.

In principle, one possible way to minimize the negative externality created by housing supply constraints in high TFP cities would be for the federal government to constraint U.S. municipalities’ ability to set land use regulations. Currently, municipalities set land use regulations in almost complete autonomy since the effect of such regulations have long been thought as only local. But if such policies have meaningful nationwide effects, then the adoption of federal standard intended to limit negative externalities may be in the aggregate interest.

An alternative is the development of public transportation that link local labor markets characterized by high productivity and high nominal wages to local labor markets characterized by low nominal wages. For example, a possible benefit of high speed train currently under construction in California is to connect low-wage cities in California’s Central Valley -- Sacramento, Stockton, Modesto, Fresno -- to high productivity jobs in the San Francisco Bay Area. This could allow the labor supply to the San Francisco economy to increase overnight without changing San Francisco housing supply constraints. An extreme example is the London metropolitan area. A vast network of trains and buses allows residents of many cities in Southern England – including far away cities like Reading, Brighton and Bristol-- to commute to high TFP employers located in downtown London. Another example is the Tokyo metropolitan area. While London and Tokyo wages are significantly above the UK and Japan averages, they would arguably be even higher in the absence of these rich transportation networks. Our argument suggests that UK and Japan GDP are significantly larger due to the transportation network.

References

ABAG (2013), "Bay Area Plan."

Ahlfeldt, Redding, Sturm and Wol, "The Economics of Density: Evidence from the Berlin Wall", *Econometrica*, 2014.

Albouy, David, "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas", mimeo, 2012.

Atkeson, Andrew Aubhik Khan and Lee Ohanian "Are Data on Industry Evolution and Job Turnover Relevant for Macroeconomics?", *Carnegie Rochester Conference Series on Public Policy*, vol. 44, June (1996), pp. 215-250.

Au Chun-Chung and Vernon Henderson (2006a), "Are Chinese Cities Too Small?" *Review of Economic Studies*, 73 (3): pp. 549-576.

Au Chun-Chung and Vernon Henderson "How Migration Restrictions Limit Agglomeration and Productivity in China," *Journal of Economic Development*, 2006b, 80, 350-388.

Avent, Ryan (2011), *The Gated City*, Kindle Edition.

Barro, Robert and Xavier Sala-i-Martin "Convergence across States and Regions," *Brookings Papers on Economic Activity*, no. 1, 1991.

Barro, Robert and Xavier Sala-i-Martin "Convergence," *Journal of Political Economy*, April, 1992.

Basu, Susanto & Fernald, John G, 1997. "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, vol. 105(2), pages 249-83, April.

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

Behrens, Kristian, Gilles Duraton, and Frederic Robert-Nicoud (2014), "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy* 122 (3): 507-553.

Close, Grank and David E. Shulenburger, *Industrial and Labor Relations Review*, Vol. 24, No. 4 (Jul., 1971), pp. 588-602

Davis, Morris, Francois Ortalo-Magne "Household Expenditures, Wages and Rents" *Review of Economic Dynamics* 14-2, 2011.

De La Roca, Jorge and Diego Puga, 'Learning by working in big cities', CEPR discussion paper 9243, December 2012.

Desmet, Klaus, and Esteban Rossi-Hansberg. 2013. "Urban Accounting and Welfare." *American Economic Review*, 103(6): 2296-2327.

Diamond, Rebecca, "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000", Stanford University, 2013.

- Diamond Rebecca, "Housing Supply Elasticity and Rent Extraction by State and Local Governments", Stanford University, 2014.
- Duranton, Gilles, Ejaz Ghani, Arti Grover Goswami and William Kerr, "The Misallocation of Land and Other Factors of Production in India" World Bank Working Paper, 2015.
- Eeckout, Jan, Roberto Pinheiro, and Kurt Schmidheiny (2014), "Spatial Sorting," *Journal of Political Economy* 122 (3): 554-620.
- Ganong Peter and Daniel Shoag, "Why Has Regional Convergence in the U.S. Stopped?", Harvard Kennedy School mimeo, 2013.
- Gennaioli, Nicola, Rafael LaPorta, Florencio Lopez-de-Silanes, and Andrei Shleifer. "Human Capital and Regional Development" *Quarterly Journal of Economics* 128 (1): 105-164, 2013.
- Gennaioli, Nicola, Rafael LaPorta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2014. "Growth in Regions" *Journal of Economic Growth* 19 (3): 259-309.
- Glaeser, Edward: "The Rise of the Sunbelt," *Southern Economic Journal*, 74(3) (2008): 610-643
- Glaeser, Edward, "The Triumph of the City", Penguin Books, 2011.
- Glaeser, Edward "Land Use Restrictions and Other Barriers to Growth", Cato Institute, 2014.
- Glaeser, Edward Joseph Gyourko and Raven Saks (2006), "Urban Growth and Housing Supply," *Journal of Economic Geography*, 6: 71-89.
- Glaeser, Edward Joseph Gyourko and Raven Saks (2005), "Why is Manhattan So Expensive?: Regulation and the Rise in House Prices", *Journal of Law and Economics*, 48(2): 331-370.
- Glaeser, Edward Joseph Gyourko and Raven Saks (2005), "Why Have House Prices Gone Up?", *American Economic Review*, Vol. 95, no. 2 : 329-333
- Gaubert, Cecile, mimeo, UC Berkeley, 2014.
- Glaeser, Edward and Joseph Gyourko, (2005), Urban Decline and Durable Housing, *Journal of Political Economy*, Vol 113, no 2, 345-375, 113 (2), 345 - 375.
- Henderson, Vernon "Systems of Cities in Closed and Open Economies," Regional Science and Urban Economics, 1982, 12, 325-350.
- Henderson and Ioannides "Aspects of Growth in a System of Cities," Journal of Urban Economics, 1981, 10, 117-139.
- Hirsch, Barry T., David A. Macpherson, and Wayne G. Vroman (2001), "Estimates of Union Density by State," *Monthly Labor Review*, Vol. 124, No. 7, pp. 51-55.
- Hsieh Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics* (2009) 124 (4): 1403-1448.

Karabarbounis, Loukas and Brent Nieman (2014), "The Global Decline of the Labor Share," *Quarterly Journal of Economics*, forthcoming.

Kline Patrick and Enrico Moretti, "People, Places and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs", *Annual Review of Economics*, 2014

Lewbel, Arthur and Krishna Pendakur, 2008, "Tricks with Hicks: The EASI Implicit Marshallian Demand System for Unobserved Heterogeneity and Flexible Engel Curves.", *American Economic Review*.

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

Moretti, Enrico, "The New Geography of Jobs" Houghton Mifflin Harcourt (2012).

Morris A. Davis & Francois Ortalo-Magne, 2011. "Household Expenditures, Wages, Rents," *Review of Economic Dynamics*, Elsevier for the Society for Economic Dynamics, vol. 14(2), pages 248-261, April

Redding, Stephen (2014), "Goods Trade, Factor Mobility and Welfare," Princeton University working paper.

Redding, Stephen "Economic Geography: a Review of the Theoretical and Empirical Literature", Chapter 16 in *The Palgrave Handbook of International Trade*, 2011

Redding, Stephen and Matt Turner "Transportation Costs and the Spatial Organization of Economic Activity", *Handbook of Urban and Regional Economics*, forthcoming, 2015.

Saiz, Albert 2010. "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, Vol. 125(3), pages 1253-1296, August.

Serrato, Juan Carlos Suarez and Owen Ziidar, "Who Benefits from State Corporate Tax Cuts? A Local Labor Market Approach with Heterogeneous Firms", mimeo, 2014

Data Appendix

In this appendix we describe where each variable used in the paper comes from. We begin by measuring average wages in a county or in a county-industry cell by taking the ratio of total wage bill in private sector industries and total number of workers in private sector industries using CBP data for 1964-65 (referred to as 1964) and 2008-2009 (referred to as 2009). To increase sample size and reduce measurement error, we combine 1964 with 1965 and 2008 with 2009. 1964 is the earliest year for which CBP data are available at the county-industry level. Data on total employment by county are never suppressed in the CBP. By contrast, data by county and industry are suppressed in the CBP in cases where the county-industry cell is too small to protect confidentiality. In these cases, the CBP provides not an exact figure for employment, but a range. We impute employment in these cases based on the midpoint of the range. We aggregate counties into MSA's using a crosswalk provided by the Census based on the 2000 definition of MSA.

The main strength of the CBP is a fine geographical-industry detail and the fact that data are available for as far back as 1964.²⁶ But CBP is far from ideal. The main limitation of the CBP data is that it does not provide worker level data on salaries, but only a county aggregate and therefore does not allow us to control for changes in worker composition. We augment CBP data with information on worker characteristics from the Census of Population and the ACS. Specifically, we merge 1964 CBP average wage by MSA to a vector of workers characteristics from the 1960 US Census of Population; we also merge 2009 CBP average wage by MSA to a vector of workers characteristics from the 2008 and 2009 ACS. These characteristics include: three indicators for educational attainment (high school drop-out, high school, college); indicators for race; an indicator for gender; and age. We drop all cases where education is missing. In the small number of cases where one of the components of the vector other than education is missing, we impute it based on the relevant state average.

Because the Census does not report information on union status, we augment our merged sample using information on union density by MSA from Hirsch, Macpherson, and Vroman (2001). Their data represent the percentage of each MSA nonagricultural wage and salary employees who are covered by a collective bargaining agreement. Their estimates for 1964 and 2009 are based on data from the Current Population Survey Outgoing Rotation Group (ORG) earnings files and the now discontinued BLS publication *Directory of National Unions and Employee Associations (Directory)*, which contains information reported by labor unions to the Federal Government. The exact methodology is described in Hirsch, Macpherson and Vroman (2001).²⁷

This allows us to estimate average residual wage in each MSA, defined as average wage conditional on worker characteristics. Specifically, we estimate residual wages as $W_{ic} - X_i'b$, where W is average wage in the MSA, X is the vector of average workers characteristics in the MSA and b is a vector of coefficients on workers characteristics from individual level regressions estimated on a nationwide sample in 1964 and 2009 based on CPS data. The

²⁶ Unfortunately, individual level data on employment and salary with geocodes is not available from the Census of Population on a systematic basis until 1980. A third of metro areas are identified in the 1970 Census.

²⁷ For 1964, estimates are calculated based on figures in the BLS *Directories*, scaled to a level consistent with CPS estimates using information on years in which the two sources overlap. Only state averages are estimated in 1964. Thus, in 1964 we assume assign union density to each MSA based on the state average.

coefficients for 1964 are: high-school or more .44; college or more .34; female: -1.13; non white: -.44; age: .004; union .14. The coefficients for 2009 are: high-school or more .50; college or more .51; female: -.45; non white: -.07; age: .007; union .14. Because a union identifier is not available in the 1964 CPS, the 1964 regression assumes that the coefficient on union is equal to the coefficient from 2009, which is estimated to be equal to .14.

For 2009, we can compare the wage residuals estimated our approach with those that one would obtain from individual level data. (Of course we can't do this for 1964, because we don't have micro data in that year). Appendix Figure 1 shows that while noisy, our imputed wage residuals do contain signal. The two measures have correlation .75.

In some models residual wage is defined as $W_{ic} - X_i'b_s$ where b_s is a vector of coefficients on workers characteristics from individual level regressions which is allowed to vary across states. The correlation in 2009 increases only marginally to .78.

Data on housing costs are measured as median annual rent from the 1960, 1970 US Census of Population and the 2008 and 2009 American Community Survey. For 1964, we linearly interpolate Census data between 1960 and 1970. Because rents may reflect a selected sample of housing units, in some models we use average housing prices. Data for 2009 are from individual level data from the American Community Survey. To get more precise estimate, we combine 2008 and 2009.

Our sample consists of 220 MSA's with non-missing values in 1964 and 2009. These cities account for 71.6% of US employment in 1964 and 72.8% in 2009. They account for 74.3% of US wage bill in 1964 and 76.3% in 2009. The average city employment is 144,178 in 1964 and 377,071 in 2009. Appendix Table A1 presents summary statistics.

Data on housing supply elasticities, land use regulations and land availability are from Saiz (2010). They are intended to measure variation in elasticity that arises both from political constraints and geographical constraints. In 19 cities, Saiz data are missing. In those cases, we impute elasticity based on the relevant state average.

Table 1a: Spatial Dispersion of Nominal Wages in 1964 and 2009

	Std. Deviation (1)	Interquartile Range (2)	Range (3)
Panel A			
Log Wage in 1964	.132	.163	.793
Log Wage in 2009	.205	.268	1.179
Panel B: Diff. with 9 Census Division mean			
Log Wage in 1964	.090	.132	.670
Log Wage in 2009	.198	.270	1.076
Panel C: Drop NY, San Francisco, San Jose			
Log Wage in 1964	.133	.171	.793
Log Wage in 2009	.157	.256	.842

Notes: The sample includes 220 metropolitan areas observed in both 1964 and 2009. All figures are weighted by employment in the relevant metropolitan area and year. Wage is the unconditional average wage in the metropolitan area. Panel B shows the distribution of the difference between log nominal wages and the average log nominal wage in each census division.

Table 1b: Spatial Dispersion of Residual Wages in 1964 and 2009

	Std. Deviation (1)	Interquartile Range (2)	Range (3)
Panel A			
Log Wage in 1964	.102	.123	.702
Log Wage in 2009	.185	.247	.920
Panel B: Diff. with 9 Census Division mean			
Log Wage in 1964	.095	.127	.612
Log Wage in 2009	.181	.221	.883
Panel C: Drop NY, San Francisco, San Jose			
Log Wage in 1964	.105	.138	.702
Log Wage in 2009	.140	.219	.686

Notes: The sample includes 220 metropolitan areas observed in both 1964 and 2009. All figures are weighted by employment in the relevant metropolitan area and year. Residual wage is the average wage in the metropolitan area after controlling for three levels of educational attainment (high school drop-out, high school, college); race; gender; age; and union status. Panel B shows the distribution of the difference between log nominal wages and the average log nominal wage in each census division.

Table 2: City GDP Growth and City Contribution to Aggregate Growth, by Group

	Accounting Estimates	Model-Driven Estimates
	Growth of City GDP As a Fraction of Aggregate Growth	Growth of City Contribution to Aggregate Growth As a Fraction of Aggregate Growth
	(1)	(2)
NY, San Francisco, San Jose	19.3%	6.1%
Rust Belt Cities (N=37)	-28.5%	6.1%
Southern Cities (N=86)	66.8%	42.0%
Other Large Cities (N=19)	31.4%	32.1%

Notes: Entries in column 1 are the growth of the city's GDP as a percentage of aggregate GDP growth over the period 1964-2009. Entries in column 2 are the percentage contribution of each city to aggregate growth from 1964 to 2009. We measure the contribution of a city to aggregate growth as the change in local TFP adjusted by the change in the gap between the local wage and the average wage as a share of the change in aggregate GDP. The group "Other Large Cities" includes 19 MSA with 2009 employment above 600,000 that are not in the other three groups. The sample includes 220 metropolitan areas observed in both 1964 and 2009.

Table 3: City Contribution to Aggregate Output 2009, by Group – Robustness to Different Assumptions on Technology

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline							
	$\alpha = .65$ $\eta = .25$	$\alpha = .70$ $\eta = .20$	$\alpha = .60$ $\eta = .30$	$\alpha = .70$ $\eta = .25$	$\alpha = .60$ $\eta = .35$	$\alpha = .60$ $\eta = .25$	$\alpha = .65$ $\eta = .20$	α and η vary by industry and year
NY, San Francisco, San Jose	6.1	6.1	6.1	6.1	6.1	6.1	6.1	5.8
Rust Belt Cities (N=37)	6.1	6.1	6.1	6.1	6.1	6.1	6.1	6.2
Southern Cities (N=86)	42.0	42.0	42.0	42.0	42.0	42.0	42.0	43.2
Other Large Cities (N=19)	32.1	32.1	32.1	32.1	32.1	32.1	32.1	31.7

Notes: Entries are the percentage contribution of each city to aggregate growth from 1964 to 2009. We measure the contribution of a city to aggregate growth as the change in local TFP adjusted by the change in the gap between the local wage and the average wage as a share of the change in aggregate GDP. Entries in column 1 are based on our baseline assumption on technology and are reproduced from Table 2, column 2. Entries in other columns vary assumptions on technology. The group “Other Large Cities” includes 19 MSA with 2009 employment above 600,000 that are not in the other three groups. The sample includes 220 metropolitan areas observed in both 1964 and 2009.

Table 4. Counterfactual Output– The Effect of Changes in the Spatial Dispersion of Relative Wages

	2009 Counterfactual Output	Percent Who Have Moved by 2009
	(1)	(2)
1) In All Cities	13.5%	52.5%
2) In NY, San Francisco, San Jose	13.2%	54.0%
3) In Rust Belt Cities	0.5%	8.7%
4) In Southern Cities	-0.4%	21.2%

Notes: Entries in column 1 are the percent difference between counterfactual output level in 2009 and actual output level. Entries in column 2 are the percent of workers who in the counterfactual scenario reside in a MSA different from their actual MSA of residence. The counterfactual involves setting 2009 relative wage equal to their 1964 level in selected cities. The sample includes 220 metropolitan areas observed in both 1964 and 2009.

Table 5. Counterfactual Output – The Effect of Changes in the Spatial Dispersion of Relative Wages - Partial Adjustment

	Percent Who Have Moved by 2009	2009 Counterfactual Output
	(1)	(2)
(1) (From Tab 4)	52.5%	13.5%
(2)	50%	13.2%
(3)	40%	11.8%
(4)	30%	9.4%
(5)	20%	6.5%
(6)	10%	3.4%
(7)	0	0

Notes: Entries in column 1 are the percent of workers who in the counterfactual scenario reside in a MSA different from their actual MSA of residence in 2009. Entries in column 2 are the percent difference between counterfactual output level in 2009 and actual output level. Row 1 reproduces Table 4, row 1. We scale partial adjustment based on the fraction of movers. For example, row (2) shows counterfactual output gains if 2009 relative wages were set so that 50% of workers relocate to a different MSA. The counterfactual involves setting 2009 relative wage equal to their 1964 level in all cities. The sample includes 220 metropolitan areas observed in both 1964 and 2009.

Table 6: Counterfactual Employment – The Effect of Changes in the Spatial Dispersion of Relative Wages

	Full Adjustment (52.5% of US Workers Move)	Partial Adjustment (20% of US Workers Move)
	Percent Change in MSA Employment (1)	Percent Change in MSA Employment (2)
Cities with Largest Increases		
NEW YORK-NEWARK, NY-NJ-PA	787.7%	179.8%
SAN JOSE, CA	522.4%	149.2%
SAN FRANCISCO, CA	509.9%	147.9%
FAYETTEVILLE-SPRINGDALE, AR	320.2%	118.1%
AUSTIN-SAN MARCOS, TX	237.7%	102.7%
City with Median Change		
SHEBOYGAN, WI	-79.7%	-32.4%
Cities with Largest Decreases		
KENOSHA, WI	-97.3%	-74.7%
MANSFIELD, OH	-97.6%	-75.7%
MUNCIE, IN	-97.8%	-76.9%
GADSDEN, AL	-97.9%	-76.1%
FLINT, MI	-97.9%	-77.4%
SHARON, PA	-98.1%	-78.3%

Note: Entries represents the percent difference between counterfactual employment and actual employment. In column 1, counterfactual employment is computed setting 2009 relative wage to 1964 levels in all cities (Panel B, first row of Table 4). In column 2, counterfactual employment is computed moving 2009 relative wage toward their 1964 levels in all cities up to the point where 20% of U.S. workers change MSA (row 5 of Table 5). The sample includes 220 metropolitan areas observed in both 1964 and 2009.

Table 7: Counterfactual Output – The Effect of Changes in Amenities

	2009 Counterfactual Output	Percent Who Have Moved by 2009
	(1)	(2)
1) In All Cities	1.6%	9.3%
2) In NY, San Francisco, San Jose	1.5%	3.1%
3) In Rust Belt Cities	-0.2%	0.8%
4) In Southern Cities	0.3%	3.7%

Notes: Entries in column 1 are the percent difference between counterfactual output level in 2009 and actual output level. Entries in column 2 are the percent of workers who in the counterfactual scenario reside in a MSA different from their actual MSA of residence. The counterfactual involves setting 2009 amenities are equal to their 1964 level in selected cities. The sample includes 220 metropolitan areas observed in both 1964 and 2009.

Table 8: Counterfactual Output – The Effect of Changing Housing Supply Regulations

	2009 Counterfactual Output	Percent Who Have Moved by 2009
	(1)	(2)
1) Regulations in New York, San Francisco and San Jose are set equal to regulations of the median city	9.70%	30.0%
2) Regulations in South are set equal to regulations in New York, San Francisco and San Jose	-3.0%	33.5%

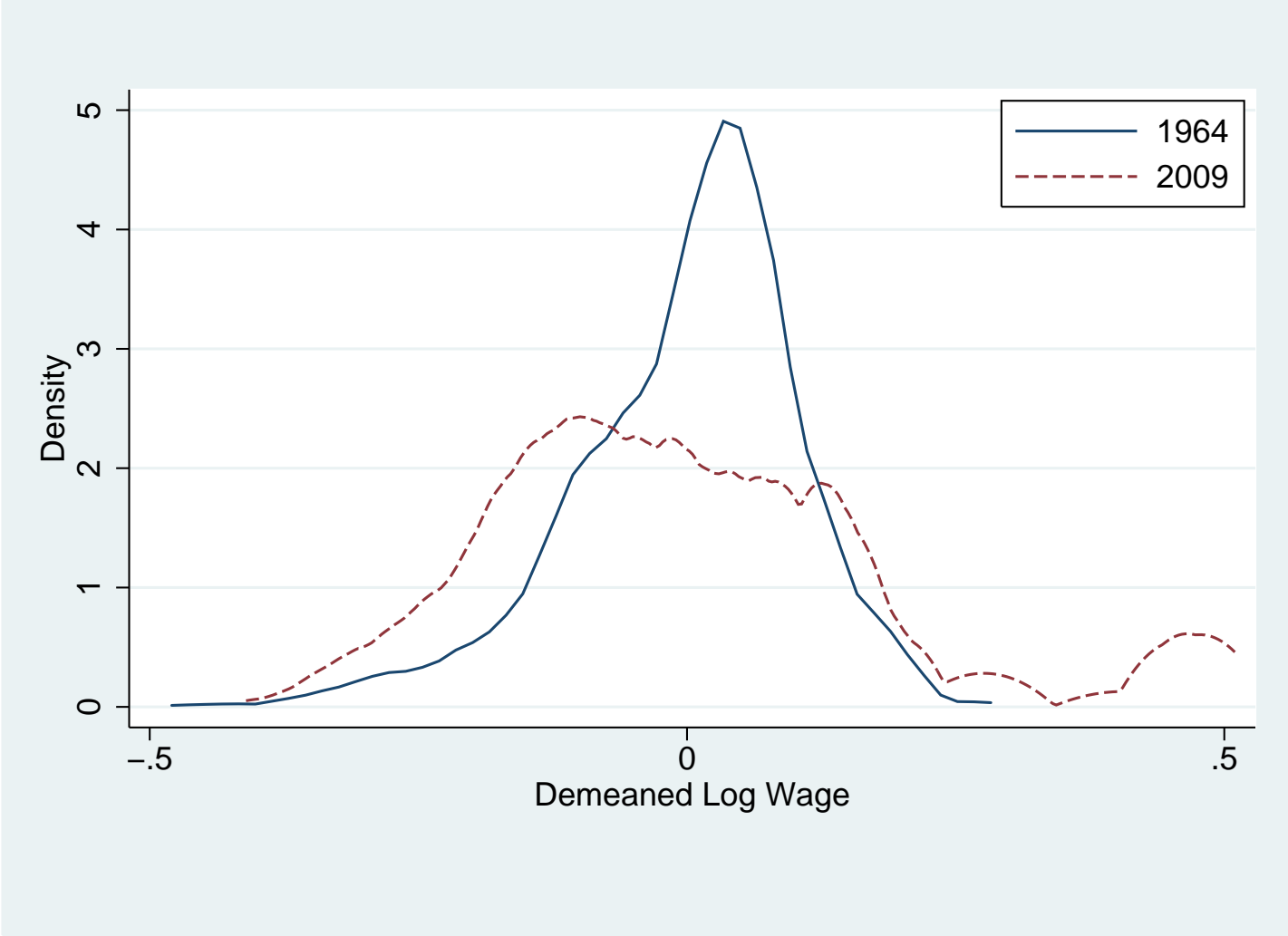
Notes: Entries in column 1 are the percent difference between counterfactual output level in 2009 and actual output level. Entries in column 2 are the percent of workers who in the counterfactual scenario reside in a MSA different from their actual MSA of residence. The counterfactual involves changing 2009 housing supply regulations in selected cities, holding land availability constant. Housing supply regulations vary at the MSA level and are measured using Saiz (2010) data, which in turn are based on the Wharton Index aggregated at the MSA level. The sample includes 220 metropolitan areas observed in both 1964 and 2009.

Figure 1a: Spatial Dispersion of Demeaned Log Nominal Wages in 1964 and 2009.



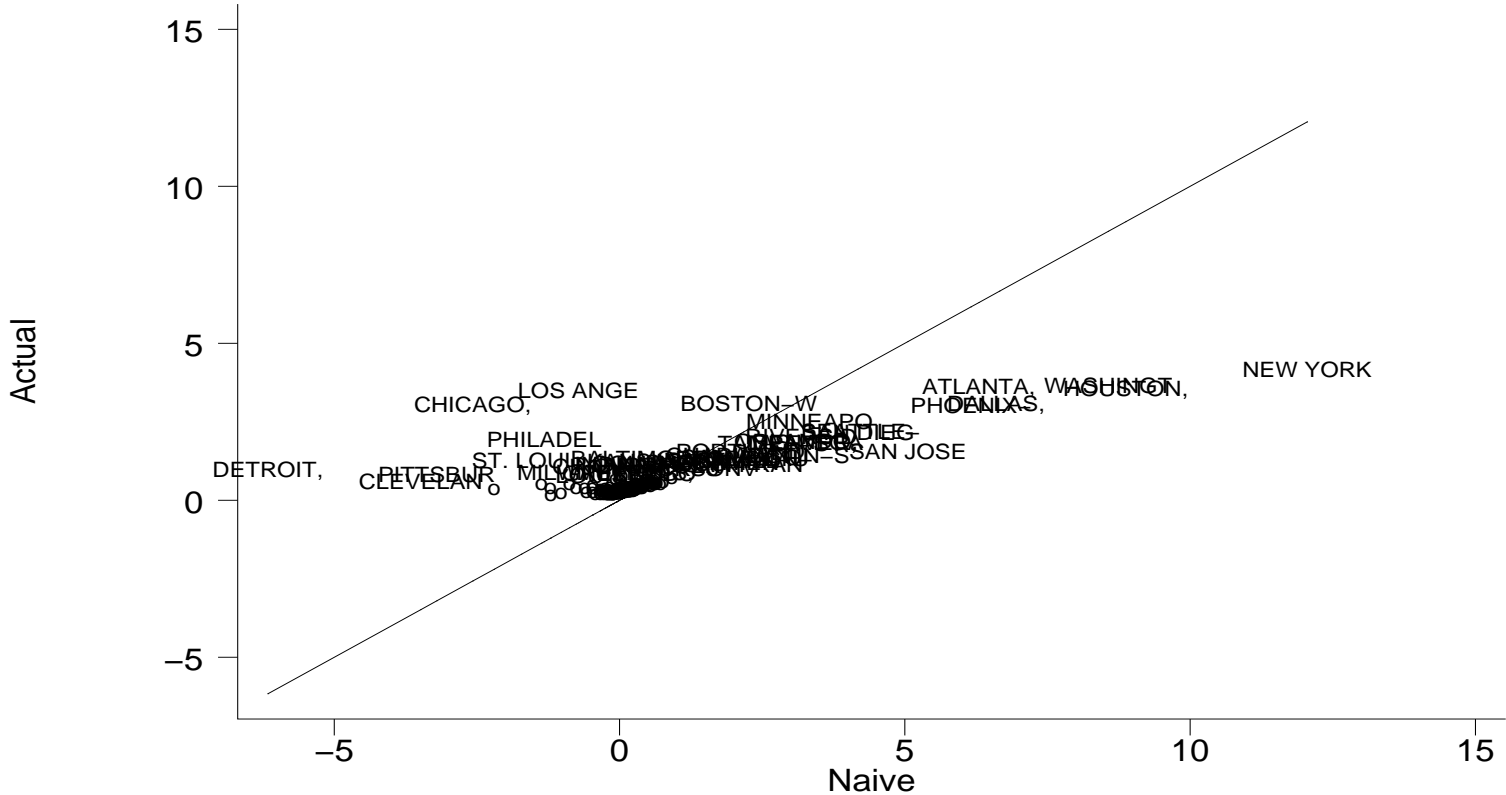
Note: The distribution is weighted by MSA employment in the relevant year.

Figure 1b: Spatial Dispersion of Demeaned Log Residual Nominal Wages in 1964 and 2009



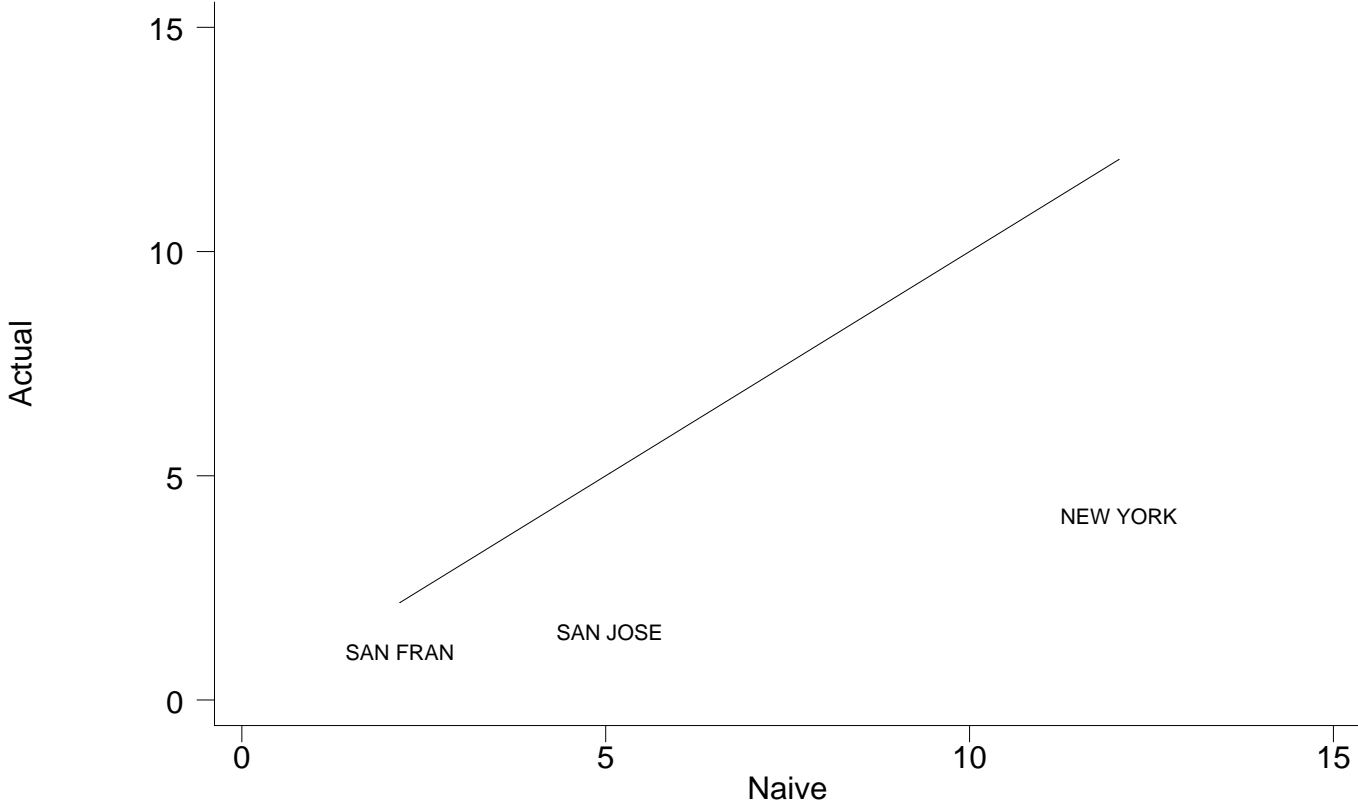
Note: The distribution is weighted by MSA employment in the relevant year.

Figure 2a: City GDP Growth and City Contribution to Aggregate Growth



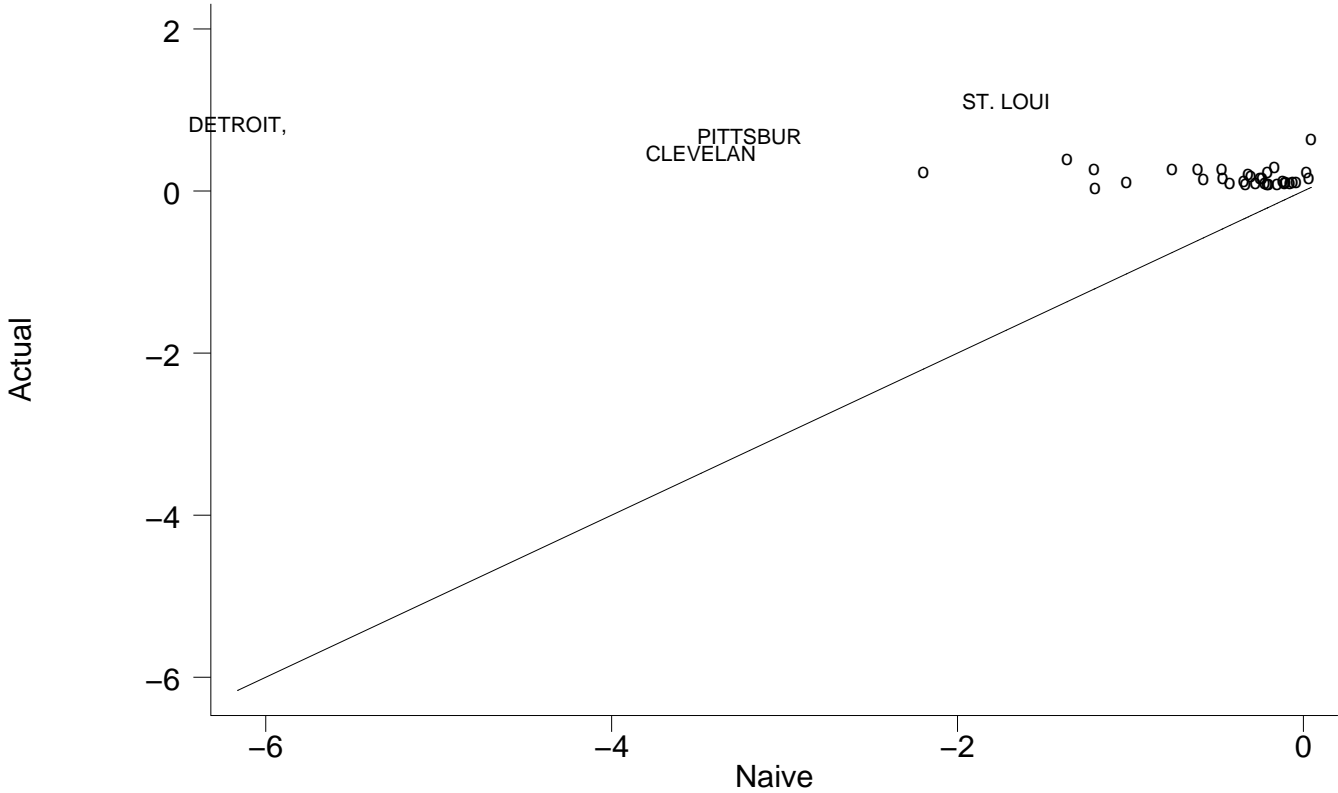
Notes: The Figure plots the percentage contribution of each city to aggregate growth from 1964 to 2009 (on the y-axis) against the growth of the city’s GDP as a percentage of aggregate GDP growth over the same period (on the x-axis). We measure the contribution of a city to aggregate growth as the change in local TFP adjusted by the change in the gap between the local wage and the average wage as a share of the change in aggregate GDP. The solid line is the 45 degree line. The sample includes 220 cities observed in 1964 and 2009.

Figure 2b: City GDP Growth and City Contribution to Aggregate Growth – New York, San Francisco, San Jose



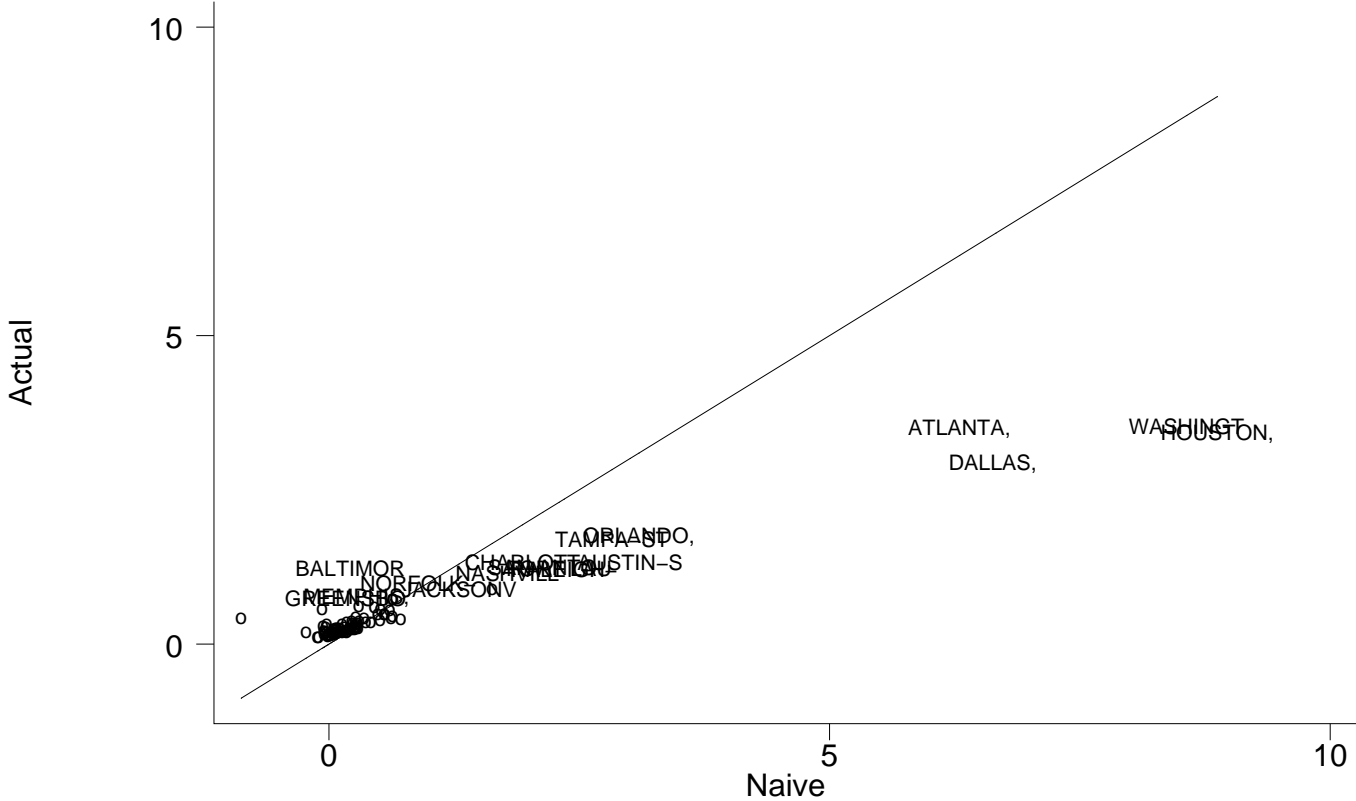
Notes: The Figure plots the percentage contribution of each city to aggregate growth from 1964 to 2009 (on the y-axis) against the growth of the city’s GDP as a percentage of aggregate GDP growth over the same period (on the x-axis). We measure the contribution of a city to aggregate growth as the change in local TFP adjusted by the change in the gap between the local wage and the average wage as a share of the change in aggregate GDP. The solid line is the 45 degree line.

Figure 2c: City GDP Growth and City Contribution to Aggregate Growth – Rust Belt Cities



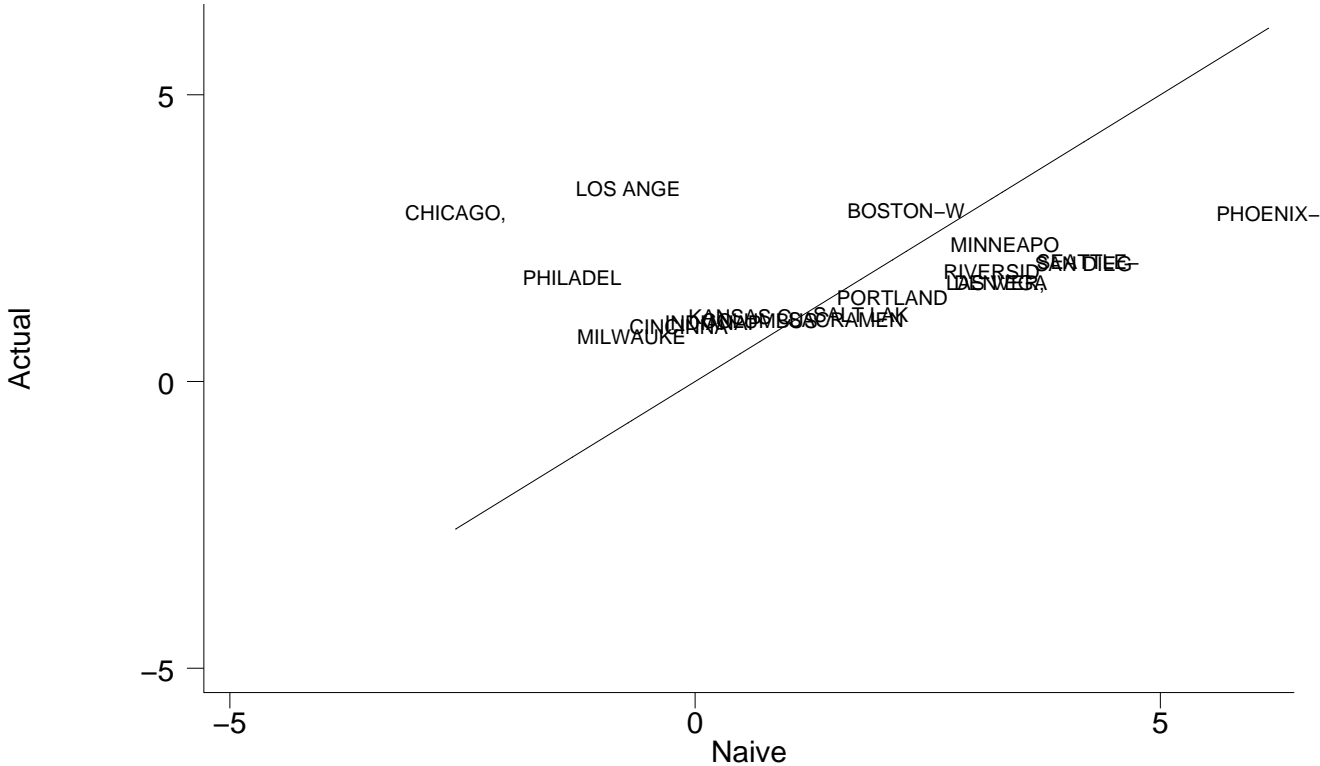
Notes: The Figure plots the percentage contribution of each city to aggregate growth from 1964 to 2009 (on the y-axis) against the growth of the city’s GDP as a percentage of aggregate GDP growth over the same period (on the x-axis). We measure the contribution of a city to aggregate growth as the change in local TFP adjusted by the change in the gap between the local wage and the average wage as a share of the change in aggregate GDP. The solid line is the 45 degree line.

Figure 2d: City GDP Growth and City Contribution to Aggregate Growth –Southern Cities



Notes: The Figure plots the percentage contribution of each city to aggregate growth from 1964 to 2009 (on the y-axis) against the growth of the city’s GDP as a percentage of aggregate GDP growth over the same period (on the x-axis). We measure the contribution of a city to aggregate growth as the change in local TFP adjusted by the change in the gap between the local wage and the average wage as a share of the change in aggregate GDP. The solid line is the 45 degree line.

Figure 2e: City GDP Growth and City Contribution to Aggregate Growth – Other Large Cities



Notes: The Figure plots the percentage contribution of each city to aggregate growth from 1964 to 2009 (on the y-axis) against the growth of the city’s GDP as a percentage of aggregate GDP growth over the same period (on the x-axis). We measure the contribution of a city to aggregate growth as the change in local TFP adjusted by the change in the gap between the local wage and the average wage as a share of the change in aggregate GDP. The solid line is the 45 degree line. This group, called “Other Large Cities” includes 19 MSA with 2009 employment above 600,000 that are not in the other three groups.

Appendix Table A1: Summary Statistics

	1964 Average (1)	2009 Average (2)
Average Annual Salary – Private Sector Workers	25,538 (3,868)	29,018 (5,278)
Average Annual Rent	4,770 (932)	6,553 (1826)
Private Sector Employment	144,178 (294,016)	377,071 (604,448)
Private Sector Wage Bill (billion)	4.04 (8.95)	13.04 (25.5)
High School Drop Out	0.59 (0.11)	0.10 (.05)
High School or More	0.40 (0.08)	0.90 (0.04)
College or More	0.07 (0.02)	0.26 (0.07)
Hispanic	0.03 (0.05)	0.10 (0.10)
Non White	.09 (0.11)	0.22 (0.15)
Age	28.1 (3.3)	39.9 (0.9)
Female	0.51 (0.01)	0.51 (0.01)
Union	0.26 (0.12)	0.11 (.06)
Number of Cities	220	220

Note: The unit of analysis is a MSA. The sample includes 220 metropolitan areas observed in both 1964 and 2009. All monetary figures are in 2000 dollars.

Appendix Table A2: Spatial Dispersion of Cost of Housing in 1964 and 2009.

	Std. Deviation	Interquartile Range	Range
	(1)	(2)	(3)
Panel A: Median Rent			
Log Rent in 1964	.205	.306	.975
Log Rent in 2009	.279	.427	1.380
Panel B: Median Housing Price			
Log Annual Cost in 1964	.278	.421	1.142
Log Annual Cost in 2009	.464	.691	2.093

Notes: Median housing price is annualized using a discount factor of 7.85% (Peiser and Smith, 1985). All figures are weighted by employment in the relevant metropolitan area and year.

Appendix Table A3: Robustness - The Effect of Changes in the Spatial Dispersion of Relative Wages Under Alternative Assumptions on Production Technology

	2009 Counterfactual Output (1)	Percent Who Have Moved by 2009 (2)
Baseline		
1) $\alpha = .65; \eta = .25$	13.5%	52.5%
Different Labor and Capital Shares, Same Returns to Scale		
2) $\alpha = .70; \eta = .20$	14.8%	55.9%
3) $\alpha = .60; \eta = .30$	12.2%	49.1%
Different Returns to Scale		
4) $\alpha = .70; \eta = .25$	29.9%	85.9%
5) $\alpha = .60; \eta = .35$	28.2%	83.9%
6) $\alpha = .60; \eta = .25$	7.2%	34.4%
7) $\alpha = .65; \eta = .20$	8.0%	37.0%
Technology Parameters Vary Across Industries and Years	7.4%	53.9%

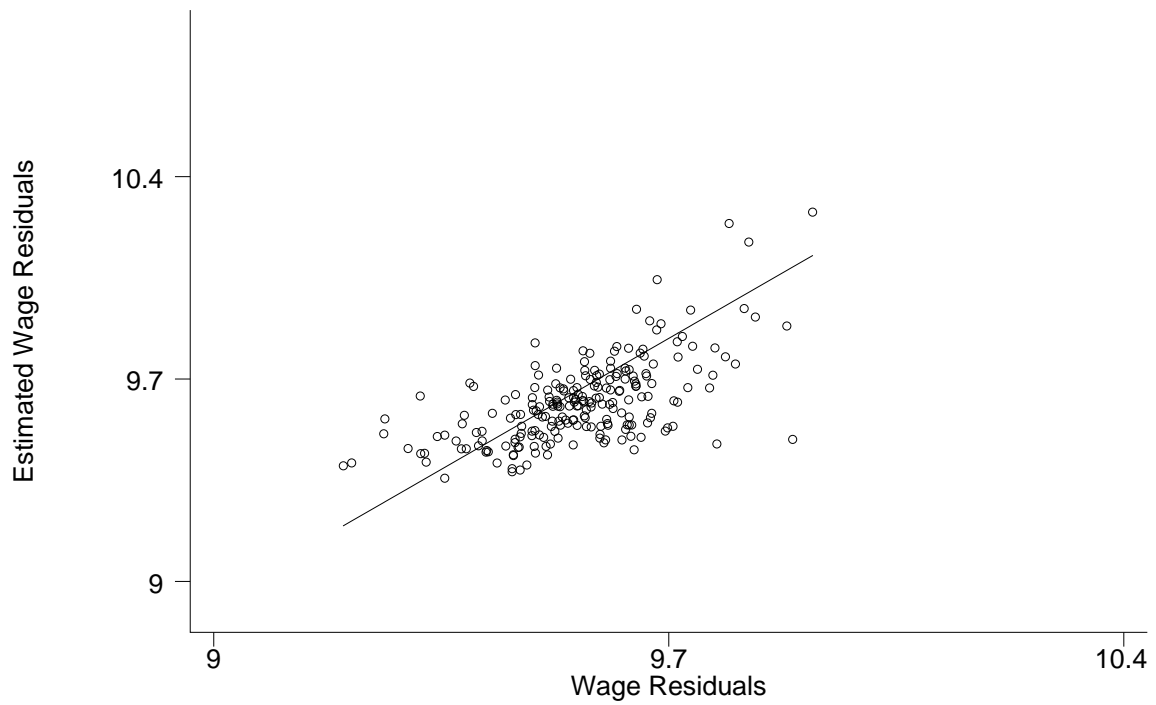
Notes Entries in column 1 are the percent difference between counterfactual output level in 2009 and actual output level. Entries in column 2 are the percent of workers who in the counterfactual scenario reside in a MSA different from their actual MSA of residence. The counterfactual involves setting 2009 relative wage equal to their 1964 level in all cities. The sample includes 220 metropolitan areas observed in both 1964 and 2009.

Appendix Table A4. Spatial Dispersion of Amenities in 1964 and 2009

	Std. Deviation	Interquartile Range	Range
	(1)	(2)	(3)
Amenities in 1964	1223.7	1737.7	6563.8
Amenities in 2009	1601.7	2304.3	7607.3

Notes: All figures are weighted by $TFP^{1/(1-\alpha-\eta)}$. The sample includes 220 metropolitan areas observed in both 1964 and 2009.

Appendix Figure A1: Estimated 2009 Average Wage Residual vs Actual 2009 Average Wage From Individual Level Data



Note: Each dot is a MSA. The x axis reports average residuals by MSA from an individual level regression based on individual level data from the Census of Manufacturers. The y axis has residuals based on CBP data used in the main analysis. The employment weighted correlation is .75.

Appendix Figure A2: Spatial Dispersion of Demeaned Log Nominal Wages in 1964 and 2009 - Unweighted

