NBER WORKING PAPER SERIES

THE WEALTH OF CITIES:
AGGLOMERATION ECONOMIES AND SPATIAL EQUILIBRIUM IN THE UNITED STATES

Edward L. Glaeser
Joshua D. Gottlieb

Working Paper 14806
http://www.nber.org/papers/w14806

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2009

Thanks to Roger Gordon and three anonymous referees for valuable comments. Glaeser thanks the Taubman Center for State and Local Government and Gottlieb thanks the Institute for Humane Studies for support. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States
Edward L. Glaeser and Joshua D. Gottlieb
NBER Working Paper No. 14806
March 2009
JEL No. D0,D00,R0,R00

ABSTRACT

Empirical research on cities starts with a spatial equilibrium condition: workers and firms are assumed to be indifferent across space. This condition implies that research on cities is different from research on countries, and that work on places within countries needs to consider population, income and housing prices simultaneously. Housing supply elasticity will determine whether urban success shows up in more people or higher incomes. Urban economists generally accept the existence of agglomeration economies, which exist when productivity rises with density, but estimating the magnitude of those economies is difficult. Some manufacturing firms cluster to reduce the costs of moving goods, but this force no longer appears to be important in driving urban success. Instead, modern cities are far more dependent on the role that density can play in speeding the flow of ideas. Finally, urban economics has some insights to offer related topics such as growth theory, national income accounts, public economics and housing prices.

Edward L. Glaeser
Department of Economics
315A Littauer Center
Harvard University
Cambridge, MA 02138
and NBER
eglaeser@harvard.edu

Joshua D. Gottlieb
Department of Economics
Littauer Center 200
Harvard University
Cambridge, MA 02138
http://www.people.fas.harvard.edu/~jdgottl/
jdgottl@fas.harvard.edu

I. Introduction

Much economic research has focused on "the wealth of nations." Many economists have tried to understand national business cycles and the higher incomes and faster income growth that occur in some countries. Yet the within-country differences in income and productivity are also quite striking. The average income per capita in 2007 in the San Francisco metropolitan area was above almost 60,000 dollars; the comparable figure for Brownsville, Texas, is under 20,000 dollars. Per capita gross metropolitan product (GMP) is more than three times higher in New York than in El Paso.[2] The differences in population density across space within countries are even more striking. As of the last census, 68 percent of Americans occupied only 1.8 percent of the country's land area.

Facts of this sort motivate the central question of urban economics: Why do cities exist? To answer this question we also must also understand why dense areas are so much more productive. Urban economists approach this question by studying, among other things, within-country variation in incomes, population density and housing prices. This allows them to treat population densities appropriately—as outcomes that are determined jointly with prices and wages. The field's central theoretical tool is the spatial equilibrium, which assumes that welfare is equalized across space—at least for marginal migrants. The spatial equilibrium concept guides urban models of housing prices and industrial agglomeration as well as empirical work on city growth and the urban wage premium.

In this paper, we review recent research on the economics of cities. We begin by presenting a version of the standard spatial equilibrium model that guides our interpretation of empirical work. The model has three core equilibrium conditions. Workers must be indifferent between locations, firms must be indifferent about hiring more workers and builders must be indifferent about supplying more housing. These three conditions provide the labor supply curve, housing supply curve and labor demand curve that collectively determine area population, wages and prices. Exogenous differences across space in productivity, amenities and the construction sector

[2] Gross Metropolitan Product is produced by the Bureau of Economic Analysis, and is meant to be comparable to Gross Domestic Product.

drive differences in density, incomes and home prices. We allow for the possibility of agglomeration economies, which exist when productivity rises with population.

We first use this model to discuss the dramatic rise of Sunbelt cities. No variable can better predict city growth over the past 50 years than January temperature, yet it is unclear *a priori* why warm places have grown so dramatically. The spatial equilibrium model enables us to assess whether this growth reflects rising Sunbelt productivity, or an increased willingness to pay to enjoy warmth, or greater housing supply in the South. Over the past 50 years, Sunbelt productivity has increased, but in the past decade incomes have fallen in warmer areas. Housing prices have also stayed low and real wages have risen in the South. These facts, interpreted using the spatial equilibrium model, imply that the recent rise of cities like Atlanta, Dallas and Houston owes more to elastic housing supply than to amenities or productivity.

The spatial equilibrium model is also needed to make sense of the dramatic connection between density and income within the U.S. (Ciccone and Hall, 1996). The standard regression of income on population makes little sense if population levels themselves endogenously increase with rising productivity. Natural factors or historical population levels are valid instruments for current population only if those instruments operate through shifts in housing supply or amenities that affect population for reasons unrelated to productivity. Despite the difficulties involved in estimating agglomeration economies, the dramatic concentration of people in high-income urban areas and the absence of obvious exogenous sources of productivity heterogeneity have lead most urban economists to believe that important agglomeration economies exist.

In Section III of this paper, we review three core theories about agglomeration economies. Cities are ultimately nothing more than proximity, so the returns to urban concentration can be seen as reductions in transport costs. One set of theories about agglomeration economies emphasizes the gains that come from reduced costs of moving goods across space (Krugman, 1991a). A second set of theories emphasizes labor market pooling and the benefits of moving people across firms (Marshall, 1890). A third set argues that cities speed the flow of ideas, which creates human capital at the individual level and facilitates innovation (Jacobs, 1968). Some of these theories emphasize the benefits that come from co-location of diverse firms; others emphasize the gains from single-industry agglomerations.

Empirical research on the sources of agglomeration economies uses information on both prices (the wages of workers) and quantities (the location of industry). For example, evidence on the co-location of industries shows that firms locate near industries that are suppliers or customers (Ellison, Glaeser, and Kerr, 2007; Kolko, 2000). The existence of human capital spillovers is suggested both by the positive correlation between individual wages and skills in the city, and by the connection between skills and city population growth.

Despite the long history of research on these questions, the field has still not reached a consensus on the relative importance of different sources of agglomeration economies. There is some evidence supporting the continued importance of transport costs for goods. Gravity models suggest that firms are drawn to areas with robust local demand. Evidence on the co-location of industries provides further support for the importance of input-output linkages in determining industrial location, but these effects are often small. However, manufacturing firms and industries with high transport costs tend to avoid dense, urban areas. These facts cast doubt on the view that cities succeed by reducing the costs of moving goods. Labor market pooling and the gains from being able to move across firms remains an important idea, but no one has managed to make the case that cities rise or fall based on this force.

A significant body of recent evidence points to the importance of skills and ideas in determining urban success. Cities with higher concentrations of skilled workers pay higher wages, and this tendency has been rising over time (Rauch, 1993). Skills predict urban growth, especially in the colder areas of the country (Glaeser, Scheinkman and Shleifer, 2005). Workers who come to cities don't receive the urban wage premium immediately, but instead experience faster wage growth rates, which suggests faster human capital accumulation in urban areas (Glaeser and Mare, 2001).

After reviewing the empirical work on the economics of urban agglomerations, we turn to the insights from urban research that might be useful for other fields. For growth and development, the evidence from cities supports the view that human capital is a particularly important source of productivity and productivity growth. Urban intellectual interactions mean that innovations are highly correlated within cities and we should expect to see significant heterogeneity in the rate of technological change across space. As a result, it may make more sense to attribute events like the English industrial revolution to the random connection of a few people than to deep-seated national characteristics.

The urban perspective emphasizes factor mobility and challenges researchers looking at sub-national data to analyze income, population and housing prices simultaneously. Running naïve income change regressions, without also considering changes in employment and housing prices, misses the fact that labor is quite mobile across places within the U.S. Housing supply elasticity will determine how much an intervention affects urban prices and quantities. These insights are essential for public economists seeking to use state-by-state variation to assess different policies and for growth economists using within-country data.

Urban economics also suggests that there are difficulties with using aggregate data to understand national inequality and income levels. Since higher-income people live disproportionately in high-income, high-cost areas, while lower-income people live disproportionately in low-cost areas, a failure to correct for local prices and amenities will overstate national income inequality. In particular, we have to be wary of using expenditure-weighted mean prices to correct for the income of the median household. Those price indices will tend to reflect the high housing prices in high-cost areas even though the median household is likely to live in a cheaper area and face a lower cost index.

House prices represent the interaction of supply conditions and the individuals' desires to live and work in certain locales. Factors such as income heterogeneity across space, amenities and land use restrictions will therefore drive housing prices. This approach is quite different from the macroeconomic perspective, which emphasizes national income and interest rates. The urban perspective on housing also differs from the standard finance perspective on housing, by emphasizing that individuals begin life short housing (Sinai and Souleles, 2003). Changes in house prices do not naturally represent an increase in national wealth since the increase in asset values has been offset by an increase in price of securing a basic necessity.

Finally, the urban emphasis on mobility implies that local poverty is more likely to reflect something good that an area is providing for the poor than a failure in local labor markets. Poor people are attracted to big cities because they offer access to public transportation and inexpensive rental housing. Further attempts to improve the lives of poor people will tend to attract more poor people to places where other low-income people live.

We now turn to the spatial equilibrium model that unifies urban economics.

II. Metropolitan Heterogeneity and Spatial Equilibrium

Just as macroeconomics explores both differences in growth rates and differences in GDP levels across countries, urban economists wonder why some cities are rich, some cities are growing, and others are doing neither. Why have some southern cities, such as Atlanta, and some former rustbelt areas like Boston seen dramatic increases in output per capita over the last 30 years, while others, like Detroit and Flint, have declined from great industrial centers to places known more for poverty than for production? How is it possible that the gap between poor and rich areas within a single country can be over 100 percent? Most importantly, why is there such a strong connection between city size and productivity, as shown in Figure 1?

One response to these puzzles is that productivity differences represent temporary aberrations that will disappear as capital flees high cost areas and labor follows higher wages. Table 1 shows the correlation between initial income levels at the metropolitan area and ex post changes in both income and population for each decade since the 1960s.[3] We have reported the coefficients and standard errors from regressing the change in the logarithm of income and the change in the logarithm of population against the initial income level.

[3] Metropolitan areas are the standard unit of analysis for much of urban economics. These areas represent multi-county groupings defined by the U.S. Census Bureau, and are more appropriate to work with than cities because they are not defined by arbitrary political boundaries.

In every decade except the 1980s, there has been substantial income convergence. Convergence seems to have slowed down since 1980, but the tendency of incomes to grow faster in poorer places, emphasized by Barro and Sala-i-Martin (1991), continues to hold. (But measurement error in income can lead to a spurious finding of mean reversion, so these facts must be treated warily.) The second column shows the relationship between initial income in each decade and subsequent population growth. If people migrate to high income areas, then we would expect initially high incomes to predict population growth. There is a positive relationship between income and population growth in the 1960s, a negative correlation in the 1970s and no significant relationship since then. Incomes are converging, but this is not because people are moving to disproportionately to high wage areas.

Does the phenomenon of income convergence suggest that current income differences are only temporary? Figure 2 shows the 0.77 correlation between the logarithm of income per capita in 1970 and income per capita in 2000.[4] There has been some convergence since 1970, but over 30 years, rich places have stayed rich and poor places have stayed poor. This continuing income disparity has motivated urban economists to think about a spatial equilibrium where differences in per capita income and prices can persist for many decades.

The Spatial Equilibrium

The methods employed by urban and growth economists differ along one major dimension. Cross-national work rarely, if ever, assumes that welfare levels are equalized across space. After all, one goal of cross-country work is to understand how to make people in poorer countries better off. However, since the pioneering work of Mills (1967), Rosen (1979) and Roback (1982), cross-city work has almost always assumed that the free migration of workers creates a spatial equilibrium where utility levels are equalized. This assumption reflects the fact that over 40 percent of Americans change homes and around 20 percent of Americans change counties every five years.

---

[4] This correlation is substantially lower if 1960 rather than 1970 is used as the initial point. The very high degrees of income convergence over the 1960 make that decade somewhat unusual over the past forty years.

The high mobility of labor leads urban economists to assume a spatial equilibrium, where elevated New York incomes do not imply that New Yorkers are better off. Instead, welfare levels are equalized across space and high incomes are offset by negative urban attributes such as high prices or low amenities. By assuming that workers choose their locations, urban economists gain at least the possibility of explaining the large concentrations of people in urban area. We can only explain city sizes if our models allow people to move into cities.[5]

Following the standard models of urban economics, we assume people can move across places immediately and costlessly. In reality, of course, these adjustments take time and money. Blanchard and Katz (1992) quantify the process of these adjustments and find that regional shocks are largely absorbed by migration flows, and house prices take around five years to adjust. Glaeser and Gyourko (2005) argue that durable housing may cause urban responses to productivity shocks to be spread over decades. Hornbeck (2008) finds that large exogenous shocks to specific regions—specifically due to soil erosion during the Dust Bowl—were mostly absorbed over long periods by largely-permanent migration. Saks and Wozniak (2007) find that migration flows respond strongly to business cycle variables, and do so differentially for workers in different stages of their careers, and Glaeser and Redlick (2008) find that education influences the size of migration flows.

The slow migration response to local shocks does not imply that spatial equilibrium holds only over long periods. As long as house prices or rents can change quickly, the price adjustment suffices to maintain the spatial equilibrium. Glaeser et al. (1995) use a spatial equilibrium model where migration responds slowly to shocks but the spatial equilibrium is always maintained because of housing price flexibility. This leads us to ask if this is occurs in practice: Do housing costs actually move enough to equalize utility levels across space?

If anything, Glaeser and Gyourko (2007) find that there is too much housing price volatility relative to volatility in local incomes. More generally, measurement difficulties mean that it is quite difficult to reject the hypothesis that welfare levels are equalized across space. The

[5] In principle, fertility differences can also explain higher density levels in some places, but fertility differences are far too small to explain heterogeneity in area population levels.

difficulties of assessing expected housing price appreciation makes it difficult to measure expected housing costs for homeowners. The fact that only a modest percentage of the population rents, and rental units are quite different from the standard stock, makes it difficult to avoid this problem by using rental prices. Researchers who use the spatial equilibrium assumption can either feel comfortable knowing that this assumption has never been rejected empirically or uncomfortable because data limitations make it impossible to test rigorously at high temporal frequencies. It is hard to see how urban economists could test the spatial equilibrium assumption with the same degree of empirical precision used by financial economists to evaluate the no arbitrage assumption in financial markets.

American scholars often find it more natural to assume a spatial equilibrium than their European counterparts. After all, migration flows are much larger in the U.S. than in Europe, and Decressin (1993) finds that population flows in Europe respond much less to local labor market shocks. Hopefully, future work will better enable to us to determine whether Europe regions are better understand as separate economies or as places linked by the free migration of labor. There is also a tradition of using the spatial equilibrium model in developing countries. The Harris-Todaro (1970) model, where high wages in big cities are offset by high levels of unemployment is a classic example.

The presence of migrants who bring the country into spatial equilibrium requires us to interpret many city characteristics as equilibrium outcomes rather than exogenous forces. In particular, a complete urban model has at least three key area-level dependent variables: wages, population levels and housing prices. These three variables are determined by three equilibrium conditions. First, workers must be indifferent across space. This ensures that real wages, corrected for local price and amenity levels, must be equalized across metropolitan areas. Second, firms must be in equilibrium, which means that wages equal the marginal productivity of labor. Third, the housing market must be in equilibrium, which requires housing prices to equal the costs of providing housing, at least in growing markets.

The individuals' location choice implies a spatial equilibrium where identical people have the same utility level across space. Following Alonso's (1964) pioneering book, a rich literature has examined the spatial equilibrium assumption within metropolitan areas. The most studied implication of that assumption was that prices would decline with commuting costs. DiPasquale and Wheaton (1996) provide the textbook treatment of this topic. Other authors have looked at the connection between housing costs and local amenities, such as good schools (e.g. Black, 1999) or disamenities, such as airports and crime (e.g. Thaler, 1978).[6]

The primary difference between the within-city spatial equilibrium work and the research that has applies the assumption across cities, is that within cities wages are typically treated as fixed. Across cities, wages differ and higher nominal wages are typically offset by higher housing costs. Higher real wages are offset by lower amenities. The spatial equilibrium yields some counterintuitive implications; for example, high real wages in an area imply that something else is bad about the place. Rochester, Minnesota, and Springfield, Illinois, are two American metropolitan areas with extremely high real wages.

While correcting for national price levels almost always makes sense, there is considerable information in local incomes that is lost by correcting for local price levels. When we assume that firms behave competitively, the marginal product of labor will be reflected in nominal local wages—not wages corrected for local prices. That information can be lost when we correct for local prices. In addition, if amenities are constant across space, then the spatial equilibrium model predicts that real wages will also be constant. Yet some places may have high wages, and high productivity levels, that are offset by high prices. Those high wages are informative because they help us to understand the correlates of local productivity. High prices will also be informative since they yield information about supply conditions in the local housing market.

People and Prices Across Space

We now turn to a benchmark model that draws on Rosen (1979) and Roback (1982) and can serve as the basis for empirical work on cities. Like growth economists, we begin with the

6 See Glaeser (2008, ch. 2) for a recent discussion of the literature on intra-city prices and allocation of people, as well as evidence supporting the rent gradient model of Alonso (1964), Muth (1969), and Mills (1967) (which itself is summarized in Brueckner [1987]). Baum-Snow (2007b) represents a significant recent addition to this body of research.

production function: $A_t^i F(K, L)$, where A_t^i is a time-city specific productivity variable, F(.,.) is a constant returns to scale production function, K is capital and L is labor. The Cobb-Douglas production function, $A_t^i K^\alpha L^{1-\alpha}$, is particularly natural for empirical work. This can either be thought of as an aggregate production function or a firm-level production function in a world with an elastic supply of firms.

The factor inputs, K and L, can represent scalars or vectors. Given the importance that skilled workers seem to play in urban growth, it is particularly natural to divide the labor force into two types of labor, but here we will treat labor as homogeneous. We will divide capital into traded capital, denoted K_T, that is bought at a nationwide, exogenous price of r_T, and non-traded capital, denoted K_N, that is city-specific and bought at local, endogenous price of r_N. The stock of non-traded capital is fixed at $\overline{K}_N$.

We assume that total capital K, is a geometric weighted average of the two types of capital so that total output is $A_t^i K_N^{\alpha\gamma} K_T^{\alpha(1-\gamma)} L^{1-\alpha}$, where γ denotes reflects the share of non-traded capital. Non-traded capital offers diminishing returns at the city level, even when firms themselves face constant returns to scale. Firms' first order condition for labor produces a city-level labor demand equation of $\varphi A_t^i \overline{K}_N^{\alpha\gamma} L^{-\alpha\gamma} = W^{1-\alpha(1-\gamma)}$, where φdepends on constant terms, including the price of traded capital. At the city level, higher wages reflect higher productivity, more non-traded capital or fewer workers.

In the economics of growth following Solow (1956), the production function is then connected to savings and investment decisions. Occasionally, the labor force itself is connected to fertility decisions (Barro and Becker, 1989). In urban models, these more dynamic considerations are generally swept under the rug. Non-traded capital is fixed and traded capital is perfectly elastically supplied at the fixed price. More sophisticated investment decisions could be brought into urban economics, and probably should be, but these issues have generally been treated as second-order.

The utility levels of workers are assumed to equal $U(G_T, G_N, \theta_t^i)$, where G_T is the consumption of traded goods, G_N refers to non-traded goods (especially housing) and θ_t^i represents local amenities. As in Roback (1982), this can be reduced to an indirect utility function $V(Y_t^i, P_t^i, \theta_t^i)$ where Y_t^i is income in place i at time t and P_t^i is the price of non-traded goods. In a static model, the spatial equilibrium assumption means that utility levels are equal across space. In a dynamic model, the spatial equilibrium assumption generally means only the lifetime utility levels will be equalized across space, but if migration is sufficiently cheap, it also implies that utility flows are equalized across space. Holding amenities constant, this yields the prediction that: $dY_t^i = -\frac{V_P}{V_Y} dP_t^i$ where the ratio $\frac{V_P}{V_Y}$ equals the demand for the non-traded good. High income levels are offset by high prices.

Again, the Cobb-Douglas utility function is a natural way to empirically use the spatial equilibrium assumption. Under this assumption, utility can be written as : $\theta_t^i G_T^{\beta} G_N^{1-\beta}$, which will equal $\theta_t^i W_t^i \left(P_t^i\right)^{\beta-1}$ times a constant. The spatial equilibrium assumption requires this to equal U_t, the reservation utility within the country. This formulation suggests that $Log(W_t^i) = Log(U_t) + (1-\beta)Log(P_t^i) - Log(\theta_t^i)$. Figure 3 shows the relationship between the logarithm of median home prices and the logarithm of median household income across space. The coefficient is 0.34, which is quite close to the average share of expenditure on housing, or $1-\beta$.

The final critical production sector concerns the making of non-traded goods, or homes. If we are interested in a truly static model, as in Roback (1982), it is natural to follow her assumption that non-traded goods are produced like traded goods with labor, traded capital and non-traded capital. In this case, the production function might be $H_t^i F(K, L)$, where H_t^i refers to productivity in this sector. We will assume that the traded capital here is the same as the traded capital in the traded goods sector, but that non-traded goods require their own, distinct form of non-traded capital (presumably land), denoted Z_N.

If total production of the non-traded good equals $H_t^i \bar{Z}_N^{\mu\eta} K_N^{\mu(1-\eta)} L^{1-\mu}$, then total output for this good will equal a constant times $\left(\left(P_t^i\right)^{1-\mu\eta} H_t^i W^{\mu-1}\right)^{\frac{1}{\mu\eta}} \bar{Z}_N$. The total labor allocated to the production of non-traded goods is $(1-\mu)(1-\beta)$ times the total population of the city. These equations can then be used to solve for city size, city wages and the wages of non-traded goods:

(1) $Log(N_t^i) = \kappa_N + \lambda_A^N Log(A_t^i \bar{K}_N^{\alpha\gamma}) + \lambda_H^N Log(H_t^i \bar{Z}_N^{\mu\eta}) + \lambda_\theta^N Log(\theta_t^i)$

(2) $Log(W_t^i) = \kappa_W + \lambda_A^W Log(A_t^i \bar{K}_N^{\alpha\gamma}) + \lambda_H^W Log(H_t^i \bar{Z}_N^{\mu\eta}) + \lambda_\theta^W Log(\theta_t^i)$

(3) $Log(P_t^i) = \kappa_P + \lambda_A^P Log(A_t^i \bar{K}_N^{\alpha\gamma}) + \lambda_H^P Log(H_t^i \bar{Z}_N^{\mu\eta}) + \lambda_\theta^P Log(\theta_t^i)$

The values of these coefficients (except for the essentially irrelevant constant terms) are given in the first column of Table 2. These equations give us the relationship between the endogenous outcomes—population, and prices—and the exogenous variables—productivity in the traded goods sector, productivity in the non-traded goods sector, and amenities. These three static equations tell us that if an exogenous variable, denoted X_t^i, influences productivity in the traded and non-traded sectors as well as amenities, then the relationship between that variable and population, prices and wages can be used to infer the impact it has on productivity, amenities and housing supply.

Let, $\frac{\partial Log(A_t^i \bar{K}_N^{\alpha\gamma})}{\partial X_t^i} = \delta_A$, $\frac{\partial Log(H_t^i \bar{Z}_N^{\mu\eta})}{\partial X_t^i} = \delta_{H,}$ and $\frac{\partial Log(\theta_t^i)}{\partial X_t^i} = \delta_\theta$ be the marginal impact of this variable on traded-sector production, non-traded production, and amenities. If the coefficients from regressions of population, wages, and price on this variable X_t^i are $\hat{b}_N$, $\hat{b}_W$ and $\hat{b}_P$ respectively, then linear combinations of these parameter estimates provide unbiased estimates of δ_A , δ_H and δ_θ. The relationship between the exogenous variable and consumer amenities can be estimated with a linear combination of the price and wage coefficient, specifically$(1-\beta)\hat{b}_P - \hat{b}_W = \delta_\theta$ This is the way to understand the Rosen-Roback contribution. Earlier work simply looked at prices to infer the impact of location-specific attributes, like crime, and utility. The Rosen-Roback correction requires also accounting for the impact that amenities may have on income. To use this equation, we need a parameter estimate for $1-\beta$, such as 0.3, the average share of household spending on housing.

An exogenous variable that impacts traded goods productivity will increase both city size and wages. Both variables must be used to infer the productivity impact of that increase. The impact of an exogenous variable on traded good productivity, δ_A, will equal the weighted sum of the coefficients on that variable in the population and wage equations; specifically, $\alpha\gamma\hat{b}_N + (1 - \alpha + \alpha\gamma)\hat{b}_W$. If non-traded capital is a particularly small part of the production process, then wages are most important for measuring the impact of a variable on productivity. In that case, even tiny increases in productivity can create large increases in the overall size of production, and as a result a city's size provides little information about productivity levels. Again, we need parameter values to implement this equation. Labor's share in total output may be two-thirds, so we can assume α=1/3. One estimate of the share of non-traded capital in production, $\alpha\gamma$, might be 0.1.

Increases in housing sector productivity will increase population and lower both wages and prices. To identify the impact of a variable on housing sector or non-traded goods productivity, we must combine the variables connection with population, income and housing prices: $\delta_H = \mu\eta\hat{b}_N + (1 - \mu + \mu\eta)\hat{b}_W - \hat{b}_P$. The parameter, $\mu\eta$, reflects the share of non-traded capital (e.g. land) in the production of housing. Glaeser, Gyourko and Saks (2005) estimate that approximately thirty percent of housing costs are associated with land and permitting across the U.S. Gyourko and Saiz (2006) estimate that 57 percent of construction costs are associated with labor costs. This suggests values of 0.4 for labor costs and 0.3 for traded capital.

The same equations are the workhorses of dynamic work on urban change. First-differencing equations (1) through (3) gives us:

(1') $Log\left(\frac{N_{t+1}^i}{N_t^i}\right) = \kappa_{\Delta N} + \lambda_A^N Log\left(\frac{A_{t+1}^i}{A_t^i}\right) + \lambda_H^N Log\left(\frac{H_{t+1}^i}{H_t^i}\right) + \lambda_\theta^N Log\left(\frac{\theta_{t+1}^i}{\theta_t^i}\right)$

(2') $Log\left(\frac{W_{t+1}^i}{W_t^i}\right) = \kappa_{\Delta W} + \lambda_A^W Log\left(\frac{A_{t+1}^i}{A_t^i}\right) + \lambda_H^W Log\left(\frac{H_{t+1}^i}{H_t^i}\right) + \lambda_\theta^W Log\left(\frac{\theta_{t+1}^i}{\theta_t^i}\right)$

(3') $Log\left(\frac{P_{t+1}^i}{P_t^i}\right) = \kappa_{\Delta P} + \lambda_A^P Log\left(\frac{A_{t+1}^i}{A_t^i}\right) + \lambda_H^P Log\left(\frac{H_{t+1}^i}{H_t^i}\right) + \lambda_\theta^P Log\left(\frac{\theta_{t+1}^i}{\theta_t^i}\right)$

Letting $\frac{\partial Log\left(\frac{A_{t+1}^i}{A_t^i}\right)}{\partial X_t^i} = \Delta_A$, $\frac{\partial Log\left(\frac{H_{t+1}^i}{H_t^i}\right)}{\partial X_t^i} = \Delta_H$ and $\frac{\partial Log\left(\frac{\theta_{t+1}^i}{\theta_t^i}\right)}{\partial X_t^i} = \Delta_\theta$, and denoting the estimated coefficients from population change, income change and non-traded goods price changes regressions by, $\hat{b}_{\Delta N}\hat{b}_{\Delta W}$, and $\hat{b}_{\Delta W}$ respectively, $\Delta_\theta = (1-\beta)\hat{b}_{\Delta P} - \hat{b}_{\Delta W}$, $\Delta_A = \alpha\gamma\hat{b}_{\Delta N} + (1 - \alpha + \alpha\gamma)\hat{b}_{\Delta W}$, and $\Delta_H = \mu\eta\hat{b}_{\Delta N} + (1 - \mu + \mu\eta)\hat{b}_{\Delta W} - \hat{b}_{\Delta P}$. As such, essentially the same formulas that are used to transform static regression coefficients into underlying parameters can be used to transform coefficients from growth regressions into parameters of interest.

One issue with this transformation, however, is the use of housing prices to capture the prices of non-traded goods. In the model, these prices, P_t^i, should be interpreted as the flow cost of non-traded services like housing. These might be appropriately identified as rental costs, but as Glaeser and Gyourko (2007) emphasize, renters are so unlike owners that rental properties often fail to give a good sense of what the flow cost of housing is within a metropolitan area. As such, we need to consider the relationship between the stock price of housing and the flow price of housing.

To move from flow costs to the stock price of housing, we must use a dynamic equation of the form $P_t^i = (1+\mu)\rho_t^i - \frac{E(\rho_{t+1}^i)}{(1+r)}$, where ρ_t^i represents the actual price of a house, μ represents maintenance and tax costs that are approximately proportional to the house price, r represents the interest rate and $E(\rho_{t+1}^i)$ is the expected price of housing next period. Solving this equation forward implies that $\rho_t^i = \sum_{j=0}^{\infty} \frac{E(P_{t+j}^i)}{(1+r)^j(1+\mu)^{j+1}}$. If the future price of housing is expected to grow at a constant rate so that $E(P_{t+j}^i) = (1+g_P)^j P_t^i$, then $\rho_t^i = \frac{(1+r)P_t^i}{r+\mu+r\mu-g_p}$, and $Log\left(\frac{P_{t+1}^i}{P_t^i}\right) = Log\left(\frac{\rho_{t+1}^i}{\rho_t^i}\right)$. Assuming a constant expected growth rate allows us to use the prices of homes as our proxy for the user cost of housing. If the constant growth rate assumption were violated, or if rates of return changed over time, then this assumption would be problematic.

An Example: Does the Rise of Sunbelt Cities Represent Amenities or Production?

To illustrate the use of the spatial equilibrium model to understand urban change, we will use it to make sense of the growth of the Sunbelt, which is among the most striking, studied and debated trends in regional economics over the last 50 years. Over the past six years, America's fastest growing metropolitan areas are Atlanta, Dallas, Houston and Phoenix. January temperature is currently positively correlated with both metropolitan area population and the growth of metropolitan area population. Cheshire and Margini (2006) find similar results within, but not across, European nations.

If we look across metropolitan areas, the relationship between January temperature and size is:

$$\text{(4) Log(Population 2000)} = \underset{(0.2)}{12.2} + \underset{(0.005)}{0.017}\text{*Average January Temperature}$$

There are 315 observations, and standard errors are in parentheses. More attention has been paid to the connection between growth and temperature across areas. In the 1990s, this relationship was:

$$\text{(4') Log(Population 2000/Population in 1990)} = \underset{(0.14)}{0.016} + \underset{(.0004)}{0.003}\text{*Average January Temperature}$$

There are again 315 observations and the R-squared is 0.162. This growth relationship is shown in Figure 4. The rise of the Sunbelt provides us with an opportunity to illustrate how the spatial equilibrium model can differentiate between different theories of Sunbelt success.

Some authors, such as Barro and Sala-i-Martin (1991) and Caselli and Coleman (2001) have emphasized capital accumulation and structural transformation in the South. These are changes that can be interpreted as increases in the productivity variables, particularly non-traded capital. Other authors, such as Besley, Persson and Sturm (2005), Olson (1983), and Cobb (1982), also point to productivity growth but suggest that this growth resulted from improvements in Southern political institutions. An alternative literature (Borts and Stein, 1965, Meuser and Graves, 1995) has pointed to consumption amenities in the South. One version of this literature emphasizes technological changes, such as air conditioning, that were complements to warmth. A second version suggests that rising incomes in the country as a whole would lead people to sacrifice productivity to live in more pleasant areas. Still a third view is that the rise of the Sunbelt's population reflects local policies that support new construction of housing (Glaeser and Tobio, 2007).

To differentiate between these hypotheses, we can use the connection between January temperature, wage growth and price growth. We begin with cross-sectional wage regressions using microdata from the 2000 Census and then move to the growth regressions that are our primary focus. In our income regressions, shown in Table 3, we include only prime-age males (between 25 and 55) and we control for education and age. In our housing price regressions, we include a battery of housing characteristics as controls. Regression (1) shows that there is a significant negative association between January temperature and incomes. Combining the coefficients from the population regression above with the coefficient from the wage regression (-0.19), yields an overall estimate of the impact of January temperature on productivity of -0.14, suggesting that warmer places within the U.S. are still less productive (by 0.14% per degree of January temperature). This may reflect the legacy of older capital investments, or it may mean that colder places with significant population levels have omitted productivity variables that justify living in the cold.

Combining the wage coefficient with the coefficient in the price regression, column (2) in Table 3, yields an amenity estimate of 0.39, meaning that people will sacrifice 0.39% of real wages per degree Fahrenheit.[7] This result is supported by the third regression showing the impact on real wages, or $W_t^i\left(P_t^i\right)^{\beta-1}$ in the model. In this case, we estimate a coefficient on January temperature of -0.33, which gives us another estimate of the amenity value of this variable. This exercise follows the Rosen-Roback static literature on quality of life variables, where papers such as Gyourko and Tracy (1989) use real incomes to infer the utility value of different local attributes. Combining population, price and income data also suggests that there is less housing supply in places with warm Januaries, which may reflect the abundance of older housing in declining, colder metropolitan areas (Glaeser and Gyourko, 2005).

We look at the connection between January temperature and growth in the 1990s by interacting January temperature with a dummy variable that equals one for observations in 2000. We also include metropolitan area dummies and either individual or housing characteristics in the wage and housing price regressions, respectively. The overall coefficient on wages, shown in column (4), is weakly negative (-0.001). The connection between housing price growth and January

[7] We are grateful to Bill Collins for pointing out errors in an earlier version of this discussion.

temperature is significantly negative (-0.43), which suggests that amenities are actually falling in warmer places over this time period. The positive interaction between January temperature in real wage growth in regression (6) suggests a similar interpretation. A more detailed look at the data suggests that the January temperature effect on real wages combines a long-run secular rise in real wages in the South and desert regions with a cyclical decline in housing prices in California between 1990 and 2000 (Glaeser and Tobio, 2007).

Combining the coefficients from all three regressions suggests that over the 1990s, January temperature was associated with neither rising productivity nor rising amenity values. Stagnant wages and declining housing prices both suggest that amenity values were not increasing. Instead, the rise of Sunbelt cities in the 1990s seems to be related to abundant housing supply, which reflects the combination of abundant land, freeways and pro-growth permitting. Over the longer run, Sunbelt status is positively associated with both productivity growth and abundant housing supply, but not with rising amenity values (Glaeser and Tobio, 2007).

The urban growth literature has paid far too little attention to the differences in housing supply that are critical to understanding the growth of metropolitan areas like Houston and Atlanta—places with high growth levels, moderate prices and moderate incomes. At the city level, the number of homes and the number of people are essentially the same thing so that differences in housing supply elasticity across space can have a large impact on how cities respond to positive shocks. Figure 5 shows the relationship between city growth from 2000 to 2007 and house prices in 2000. While many places saw high prices along with expansion, the pattern clearly fails to hold for about 40 of the most expensive cities. These places had virtually no population growth but still had the highest house prices in the country. Simultaneously, the 20 or so places with the largest population growth had moderate prices. Since the places that expanded the most are not expensive and the places that are expensive did not expand, housing supply must differ across areas.

III. Agglomeration Economies and the Existence of Cities

The central question of why cities exist ties together almost all of urban economics. Most of the field that follows von Thunen (1825) and Marshall (1890) can be understood as an attempt to make sense of the remarkable clustering of human activity in a small number of urban areas. The spatial equilibrium model again provides the grounding for thinking about the reasons for urban concentration, which again might reflect consumer amenities, housing supply or productivity advantages. Moreover, cities might form because some places have innate advantages in these areas-- New York City's harbor increases productivity and San Diego's climate is pleasant—or because clusters of people endogenously increase amenities, housing supply or productivity.

There are certainly cases where cities have formed for consumption rather than production reasons. The early history of Los Angeles, for example is replete with examples of people, such as prosperous Midwestern retirees, who came there to enjoy the climate. Urban amenities, such as a thriving restaurant or theater scene, can occur endogenous with an influx of population. However, the spatial equilibrium model allows us to easily reject the view that consumer amenities are the primary force driving urban concentration in the U.S. If cities were driven by amenities, then real wages should be lower in big urban areas, and this is not the case. The real wage premium associated with living in big cities has declined over time (Glaeser and Gottlieb, 2006) which suggests that cities have become relatively more pleasant places to live, perhaps because of the decline in crime (Schwartz, Susin and Voicu, 2003). Yet even today, however, people require a mild wage premium to locate in big urban areas.

It is even more implausible to think that big cities exist because these areas have an innate advantage in supplying housing or because density makes it easier to build. Data on construction costs show that it costs considerably more to build vertically than horizontally (Gyourko and Saiz, 2006). Across metropolitan areas, housing prices rise substantially with city size, so the spatial equilibrium approach suggests that housing supply is more expensive in those places. There are particular places, such as the growing Sunbelt metropolises or the declining cities of the Midwest, where abundant housing supply boosts population, but this is not generally true for urban areas.

We are left with the view that cities exist because they are areas with high levels of productivity, which might occur because people come to places that are innately more productive or because density itself enhances productivity because of agglomeration economies. The strong correlation between urban size and productivity, shown in Figure 1, supports this view. Does that link reflect agglomeration economies, where size creates productivity, or heterogeneous local productivity levels which then cause agglomeration?

Many cities have undoubtedly benefited significantly from innate productivity advantages, particularly waterways. The rise of New York City is intimately connected with the strengths of its spectacular natural harbor (Albion, 1938). Every one of the 20 largest American cities in 1900 was on a major waterway, reflecting the enormous cost savings associated with moving goods over water.[8] Pittsburgh and its iron industry owes much to nearby coal mines. Chicago's stockyards benefited from proximity to the rich, black soil of Iowa (Cronon, 1990).

Davis and Weinstein (2001) examine the continuing importance of natural advantage by looking at population growth among Japanese cities that were devastated during World War II. Despite significant population losses during the war, bombed cities almost all returned to their pre-war growth paths. If agglomeration economies were very important, then these cities might have been derailed from their long run growth paths. Since significant losses in population during the war did not materially impact the growth in these areas, agglomeration economies might not be all that important relative to underlying fundamentals. However, it is quite possible that the natural advantages of these places were actually man-made improvements, such as transportation networks that either survived the war or were rebuilt. As such, it is hard to interpret these results as strong evidence for true innate advantage.

[8] An 1816 U.S. Congress report stated that it cost the same amount to move goods across the Atlantic as to take them 30 miles inland (Taylor, 1977).

The substantial decline in shipping costs over the 20^{th} century makes it hard to believe that innate productivity advantages remain terribly significant. The cost of moving a ton by rail has declined in real terms by more than 90 percent since the late 19^{th} century and the rise in trucking has been even more dramatic (Glaeser and Kohlhase, 2004). As a result, access to the Great Lakes, other water systems and raw materials like coal and corn, has become less valuable over the 20^{th} century. The growth of cities without access to major waterways has accompanied this decline. A century-long county growth regression yields:

$$(5)\ \text{Log(Population 2000/Population in 1900)} = \underset{(0.03)}{0.62} + \underset{(0.00016)}{0.00122^{*}}\ \text{Distance to River}$$

In this case, we use counties rather than metropolitan areas, since selection into the sample of metropolitan areas is itself an indicator of substantial success over this longer time period. There are 2,804 observations and the R-squared is 0.02. We define distance to a river using rivers identified by Fogel (1964) as navigable in 1890, from the Rappaport and Sachs (2003) data. The regression shows that places farther from rivers have grown more quickly, but the actual explanatory power of this variable is quite weak—possibly because water-related natural advantages might still matter for production or consumption.

Surely, there are surely some innate advantages associated with particular areas. Moreover, many places are more productive because of decades, if not centuries, of investments in productive infrastructure. Most of the U.S. literature has implicitly assumed that political forces have only a modest impact on local productivity. This assumption is probably more palatable in the U.S. than it is in many other places, particularly developing countries. Ades and Glaeser (1995), for example, argue that outside of stable democracies there is a large return to firms in being close to the corridors of power. Those returns then attract people and firms, and help explain why capital cities are so much larger in unstable or dictatorial nations. While the bulk of the literature on cross-city income differences has focused on agglomeration economies, the empirical quest to accurately measure such economies has proven to be quite difficult, as the next section illustrates.

Agglomeration in Productivity

Agglomeration economies are, at their root, advantages that come from reducing transportation costs. After all, urban density is just the absence of physical space between people and firms. Agglomeration economies can exist because of reduced transportation costs for goods: input suppliers and customers save on those costs if they locate near one another. Agglomeration economies can exist because of reduced transportation costs for people: labor markets may be more efficient in urban areas and service providers may find it easier to cater to their customers. Finally, agglomeration economies can exist because of easier transmission of ideas: cities may thrive because they facilitate the flow of knowledge across people and enterprises.

The most natural adjustment to the model is to assume that productivity is also a function of city size: $A_t^i = a_t^i \left(N_t^i\right)^{\omega}$, where a_t^i is a parameter and N_t^i is current population. This adjustment is a reduced-form method of incorporating the agglomeration economies that are explicitly derived in the models of the New Economic Geography (NEG) that follow Krugman (1991a). In those models, transportation costs are explicitly modeled, and as a result some places are more productive than others. Here, we just assume that a simple function captures those agglomerating forces. There is no conflict between the spatial equilibrium models that follow Rosen and Roback and the New Economic Geography models that follow Krugman, which also assume a spatial equilibrium. In a sense, the NEG models are just more complete in there derivation of productivity differences.

With this adjustment, equations (1), (2), and (3), and equations (1'), (2') and (3') continue to hold, but the parameters in those equations need to be interpreted somewhat differently. The second column of Table 2 gives the value of these equations' parameters when there are these agglomeration economies. As long as $\frac{(1-\alpha)(1-\beta)\eta\mu+(\mu+\beta-\mu\beta)\alpha\gamma}{\mu+\beta-\mu\beta-(1-\beta)\eta\mu} > \omega$, the system will continue to be stable.

Since rising productivity attracts more people which further increases productivity, agglomeration economies act as a multiplier that increases the relationship between exogenous productivity-enhancing factors and population, wages and housing prices. Agglomeration economies strengthen the positive relationship between amenities and housing supply and population for the same reason. The negative impact that amenities and housing supply have on prices and wages can be increased, muted or reversed by agglomeration economies. If $\omega > \alpha\gamma$, for example, then rising housing supply and rising amenities actually increase wages and housing prices, because higher population levels increase productivity through agglomeration economies.

One can begin to make such measurements by regressing either income or productivity on city size or density. Ciccone and Hall (1996), for example, show a strikingly powerful connection between density and productivity across states. Combes, Duranton, Gobillon and Roux (2009) perform similar regressions using French data. As Figure 1 shows, bigger metropolitan areas, which are also denser, are more productive.[9]

Regressions of this type, however, run counter to the whole spirit of the spatial equilibrium approach, which requires us to treat area population as an endogenous variable.[10] Since migration lies at the very heart of the urban model, even those urban economists who believe in agglomeration economies most fervently find it hard to treat city size as an independent variable. Ciccone and Hall (1996) and Combes et al. (2009) are well aware of this endogeneity problem and they use an instrumental variables strategy. But to make sense of an instrumental variables approach to agglomeration economies, we must specify a full spatial model that incorporates those economies.[11]

The modified coefficients for equations (1), (2) and (3) tell us how to interpret instrumental variables estimates. We cannot interpret any empirical relationship between population and income unless we know whether the source of variation across communities is productivity,

---

9 The relationship between density and per capita Gross Metropolitan Product across metropolitan areas is slightly less than the correlation between area population and GMP. In a multivariate regression, both variables are significant.

10 Greenstone, Hornbeck, and Moretti (2007) find positive effects of winning a million dollar plant on area productivity. By looking at high enough frequencies, their work can be seen as an attempt to estimate agglomeration externalities before population has the opportunity to fully adjust.

11 The problems of relating city size to contemporaneous incomes or housing prices become less severe when we look at the relationship between initial income and later population or income changes. In this case, the coefficients can be interpreted as the impact of initial size and later changes in amenities, productivity and housing supply. Glaeser, Scheinkman and Shleifer (1995) and Eaton and Eckstein (1997) both find no connection between initial city size and later growth. Gabaix (1999) then shows that this non-relationship (Gibrat's law) can explain Zipf's law, and documents that this relationship holds across countries and time periods.

amenities or housing supply. If we were sure that productivity were constant across space, so that the variation was coming from housing supply or amenities, then the ordinary least squares coefficient estimated from regressing the log of income on the log of population would equal $(\omega - \alpha\gamma)/(1 - \alpha + \alpha\gamma)$. That is the same coefficient that would be estimated in an instrumental variables regression using instruments that capture housing supply or amenities but not productivity. Alternatively, if an exogenous productivity shock is used to instrument for population in a regression of income on population, the procedure will estimate λ_A^W/λ_A^N, or $\mu\eta(1-\beta)/(\beta + \mu(1-\beta)(1-\lambda))$.

The instrumental variable estimate will never yield an unbiased estimate of ω, the true treatment effect of population on productivity. If heterogeneous productivity provides the source of cross-area variation, then this parameter does not even enter into the estimated coefficient. If amenities and housing supply generate cross-area heterogeneity, then the agglomeration effect does enter into the estimated coefficient, but it is scaled down by $\alpha\gamma$, the share of production associated with non-traded capital, which creates congestion in the production process. The effect is also scaled up by $1 - \alpha + \alpha\gamma$, the share of production associated with labor plus non-traded capital.

Table 4 shows relatively standard ways of estimating agglomeration economies. In the first regression, we show the 0.04 elasticity of income with respect to city size. The second regression uses the population of the agglomeration in 1880 as an instrument for population today. This procedure mimics Ciccone and Hall's use of long-standing historical variables to instrument for city population. But how do we interpret historical population levels in light of the model?

If historical population is correlated with the productivity level of the area it will not give us an estimate of the agglomeration effect. Ciccone and Hall (1996) suggest that the U.S. economy has changed so much that it is unlikely that the variables relevant for productivity in the mid-19th century still matter today. This may well be the case, but there is still the possibility that older areas have more non-traded physical capital or intangible assets that have been built up over time. If those factors make the region more productive, then the instrumental variables estimates will not tell us about agglomeration economies. Keeping these concerns in mind, we find that this procedure produces an elasticity estimate of 0.08.

Regression (3) shows the alternative approach of using weather variables. This mimics Rosenthal and Strange (2003) and Combes et al. (2009) who use geological variables to instrument for current density levels. We use January and July temperature, precipitation, longitude and latitude as instruments. In this case, the estimated elasticity is 0.04, which is the same as the ordinary least squares estimate. If $\alpha\gamma$ equals 0.1 and $1 - \alpha + \alpha\gamma$ equals 0.76, and if these instruments reflect amenities or housing supply, then the estimated coefficient suggests a value of 0.13 for ω. We are not particularly confident that these, or any current instruments, are orthogonal to productivity, so any interpretation is hopeful at best.

In regressions (4)-(6), we look at housing prices and city sizes. Regression (4) shows that housing prices are higher in big cities, but this ordinary least squares coefficient is difficult to interpret. If we use productivity variables to instrument for city size, the instrumental variables estimate should equal $\mu\eta/(\beta + \mu(1-\beta)(1-\lambda))$. If we use amenity variables to instrument for city size, the coefficient should equal $((1-\alpha+\omega)\eta\mu - (1-\mu)(\alpha\gamma-\omega))/(1-\alpha+\alpha\gamma)$. If housing supply variables are used to instrument for city size, then the estimated coefficient should be $-(\alpha\gamma-\omega)/((1-\alpha+\alpha\gamma)(1-\beta))$. Again, the nature of the exogenous source of variation is critical in interpreting the coefficients. City size is positively associated with housing prices in all three specifications, but the coefficient is by far the highest in regression (6), which may reflect the fact that the geography variables are most highly correlated with amenity variables and less highly correlated with housing supply.

The remaining regressions in Table 4 look at real wages. In regressions (7)-(9) we find that real wages are not significantly associated with larger city size in OLS or using historical population as an instrument. In the amenity-driven IV specification, however, there is a significant negative effect of population on real wages. This suggests that amenities may be higher in bigger cities today than they have been in the past.

The existence of industrial clusters also seems to suggest that agglomeration economies are important (Krugman, 1991b). There is a long tradition of examining such clusters to better understand agglomeration economies (Hoover, 1948, Fuchs, 1957). However, the same basic identification problem that plagues efforts to infer agglomeration economies by looking at city size-city income relationships also troubles research on industrial concentration. If natural advantage is heterogeneous across space, then we should expect to see industries clustering in particular locales even without agglomeration economies (Ellison and Glaeser, 1997).[12] In principle, if industries are sufficiently footloose, then even a tiny edge in a particular locale can produce extreme industrial concentration. This tendency will be even greater if the industry is concentrated in a small number of manufacturing plants.

One way to address the fact that industrial concentration can easily reflect both omitted natural advantages and agglomeration economies is to try to correct for observables sources of natural advantage. Ellison and Glaeser (1999) do that and find that controlling for an extreme large number of natural advantage variables reduces the average level of geographic concentration by twenty percent. The share of industries with extremely high levels of geographic concentration declines from 12.8 percent to 11.1 when controlling for a rich set of location characteristics. We cannot be sure that omitted variables are not significantly more important than the observed variables, but the relatively modest ability of observed variables to explain geographic clustering suggests that this clustering may reflect agglomeration economies of different forms.

While estimating agglomeration economies is difficult, there remains a robust consensus among urban economists that such economies exist. The concentration of industries, the concentration of people in cities, and the higher urban incomes and productivity levels all point to the existence of these benefits from city size. Yet they do not tell us exactly why they exist, the question to which we now turn.

People cluster in cities to be close to something. At their heart, agglomeration economies are simply reductions in transport costs for goods, people and ideas. We start by discussing

---

[12] Ellison and Glaeser (1997) provide an index of agglomeration using discrete spatial data; Duranton and Overman (2005) find statistically significant agglomeration using continuous geographic information.

transport costs for goods, and then discuss the role of cities as labor markets and places of idea transmissions.

Transport Costs and Agglomeration

Smith (1776), von Thunen (1825) and Marshall (1890) all discuss the role that cities can play in reducing transport costs. However Krugman (1991a) is appropriately given credit for crafting an internally consistent model where spatial concentration reflects the desire to cut shipping costs. In Krugman's initial formulation, the benefits of agglomeration come from reducing the costs of moving goods across space. That paper and the ensuing literature, some of which is described in Fujita, Krugman and Venables (1999), remind us that fixed costs as well as transport costs are needed to explain agglomeration. Without some scale economies, firms could divide into arbitrarily small sub-components and spread themselves throughout the hinterland. Fixed costs ensure that businesses will only have a moderate number of facilities and then transport costs push those facilities close to their suppliers and customers.

Urban history is replete with examples of industries locating in cities to reduce transport costs. As Zipf (1949) noted, transport costs can cause industries to locate at the point where inputs are first produced, at the point of final consumption, or at a central spot in between. In Pittsburgh, a city formed around coal, a basic input into steel production. Meatpacking located in Chicago because cows and pigs came through that city as they traveled from the agrarian west to eastern markets. The vast stockyards exploited economies of scale. As scale economies declined, the stockyards left Chicago and moved upstream into the agricultural hinterland. In this case, the transport cost advantages came from Chicago's role as the hub of a major transportation network, not from selling to the city's residents. By contrast, New York's dominant industries in the 19th century—sugar refining, garment manufacturing and printing and publishing—all sold much to New Yorkers. These industries certainly took advantage of inputs coming through the city's harbor (raw sugar, textiles and pirated English novels), and they also were big exporters. Yet in this case, the access to local demand also helped drive their urban locations.

A number of papers seem to show the importance of market demand or supply for productivity or industrial location. Hanson (2005) connects employment and wages with the demand in neighboring areas, or market potential. He estimates a variant of the Krugman (1991a) model and finds support for strong input-output linkages. One problem with this estimation is the endogeneity of demand in neighboring areas. Head and Mayer (2004) address this issue by looking at the location of Japanese affiliates in Europe, and find that location choices are quite correlated with pre-existing market demand. Davis and Weinstein (2005) look at Japanese data and find that productivity rises with both market demand and input supply, though they have an endogeneity problem because the location of production may induce consumers to locate in a particular area.

Cross-industry co-location patterns provide another source of evidence on the importance of input-output linkages. Ellison, Glaeser and Kerr (2007) find a moderate tendency of manufacturing firms to locate near industries that are either their suppliers or purchasers. The endogeneity problem here is that industries end up buying from other sectors that are geographically close. The authors address this problem by looking at purchasing patterns in the U.K. and in areas where industries are not co-located.[13]

Transport costs undoubtedly remain important for many industries. Figure 6 shows the connection across industries between average shipment length against the logarithm of value per ton, which measures how heavy the goods are per unit of value. Those goods that weight the most are shipped the shortest distances, which confirms the view that transport costs still matter, at least for shipping patterns. However, it isn't clear that high transport costs increase either geographic concentration or urbanization. If we correlate the Ellison and Glaeser (1997) geographic concentration index with the log of value per ton across industries, there is no significant correlation.

If we correlate the log of value per ton with the share of the industry's employment in metropolitan areas, we find a significant positive relationship, shown in Figure 7. High transport

[13] Henderson and Ono (2008) specifically distinguish between the location of a firm's headquarters and its manufacturing facilities. This enables them to analyze the supply of business services to manufacturing firms, and they find it to be an important determinant of headquarter location.

cost industries locate away from urban areas. This fact pushes us away from the view that cities exist to reduce transport costs for hard-to-ship products. The fact that manufacturing firms generally locate outside of big cities further suggests that cities have lost their historic comparative advantage at moving goods cheaply.

Cities today are much more likely to specialize in business services, and Kolko (2000) finds that business service firms are much more likely to locate near sectors that are abundant in potential customers or input suppliers. Indeed, it is natural to think that transport costs are more important for service firms where output typically involves face-to-face contact. Reducing the transport cost of purchasing inputs may still be an important element in urban economies, but the important inputs are now management consultants rather than iron ores.

The one missing element in applying a theory designed to explain the concentration of manufacturing to business services is an understanding of the fixed costs in service industries. In the case of manufacturing there are scale economies in large plants, like the stockyards and sugar refineries, that are lost if production is spread through the hinterland. What is the equivalent force that applies to business services? One hypothesis is that the benefits of specialization create increasing returns in business services. In a large city, with abundant clients, it is possible to specialize in a narrow area, which will improve quality and reduce the need for general training (Becker and Murphy, 1992). Even Adam Smith (1776) wrote that division of labor is limited by the extent of the market. In large markets, business service providers can specialize more completely, reaping all of the associated benefits.

Access to Workers and Dense Labor Markets

The core idea of labor market-based agglomeration economies can be described as firms looking for workers and workers looking for firms. While this is surely true, without more elaboration the argument doesn't predict large agglomerations—just that one firm and some workers will locate near one another. Alfred Marshall (1890) argued that larger agglomerations could result if there were shocks to firm demand or productivity. In that case, when multiple firms locate near one another, workers can move from firms that have experienced negative shocks to firms that have received positive shocks. If workers are risk averse, this increases average productivity through the process known as statistical returns to scale. Krugman (1991b) offers an elegant model of this concept.

Similar gains to labor market agglomerations occur if there is uncertainty about match quality between worker and firm. If firms are in splendid isolation, then workers will be stuck with their first employer. If there are many firms locating near one another, there is opportunity for workers to hop from job to job in order to find the best match for their talents and interests (Helsley and Strange, 1990). Strange, Hejazi and Tang (2006) provide a general model linking uncertainty and the gains from agglomeration that come from statistical returns to scale. Changes in preferences over the lifecycle, such as a desire to change work patterns for parenting, create advantages from agglomeration even without uncertainty.

There are numerous cases in which dense agglomerations provide extremely well-functioning labor markets. For example, Wall Street's dense labor market concentration allows workers to hop from firm to firm.[14] The Connecticut hedge fund industry took advantage of the already dense agglomeration of financial services workers in the New York area and located close to many of their suburban homes. Likewise, Silicon Valley provides plenty of opportunity for job-hopping engineers to move firms (Saxenian, 1994; Fallick, Fleischman and Rebitzer, 2006). The presence of so many alternative employers also may induce workers to take on riskier jobs because they know that there will be other employment opportunities if a venture goes bust.

Yet there has been relatively little empirical work documenting the importance of labor market pooling. Diamond and Simon (1990) show that workers in more specialized cities, who face greater employment risk, are compensated for that risk with higher wages. Costa and Kahn (2000) argue that highly skilled couples locate in big cities because of the benefits of thick labor markets. Ellison, Glaeser and Kerr (2007) find that industries locate near other industries that use the same type of workers. Overman and Puga (2007) use U.K. data and find that industries with more idiosyncratic risk are more likely to cluster near one another. This evidence is

---

[14] Rosenthal and Strange (2008a) present evidence showing that dense urban labor markets also increase hours worked for young professionals, possibly because they are trying to compete with their peers.

suggestive, but more work is needed to establish this as a major force driving either industrial location or urbanization.

The theory of labor market pooling suggests that the gains from co-location will be highest for firms that use the same types of workers but are subject to different labor market shocks. This type of argument suggests that in some cases, agglomeration economies may come from locating near firms that do similar things and sometimes the gains come from locating near a wider range of industries. Following Henderson (1988), agglomeration economies that operate within industries are often called localization economies while agglomeration economies that work across industries are referred to as urbanization economies. Most of the interest in these different categories has come from researchers focused on the knowledge-related agglomeration economies.

Cities and Ideas

Over the past 15 years, following the "new growth economics" of Romer (1986) and Lucas (1988), urban economists such as Glaeser, Kallal, Scheinkman and Shleifer (1992) and Rauch (1993) have increasingly focused on the role of location-specific ideas and human capital. Idea-based agglomeration economies result when ideas move imperfectly over space, as suggested by Marshall (1890), Jacobs (1968), Helsley and Strange (2004) and many others. The key piece of empirical evidence supporting this claim comes from Jaffe, Trajtenberg and Henderson (1993) who show that patents are more likely to cite previous patents that are geographically proximate.

There are several ways in which cities become more productive as centers of idea transmission. Faster intellectual flows in cities make firms more productive at any given point in time. Up-to-date information is a direct input into some industries, including journalism and finance. The spread of ideas in cities may also increase the rate of technological change in those areas. Anecdotal support for this view is given by Saxenian (1994) who discusses the communication of new ideas across firms in Silicon Valley in social settings, like the Wagon Wheel restaurant. Audretsch and Feldmann (1996) show that commercial innovations are concentrated in urban areas.

Duranton and Puga (1998) present a version of this hypothesis suggesting that cities are "nurseries" for new ideas. In their model, mature industries then flee cities for the lower production costs of non-urban locales. The Ford Motor Company is a good example of a firm that had its most innovative stages in central city Detroit, where it could easily acquire inputs for its prototypes. Ford then moved to suburban River Rouge to lower costs when its product, the Model T, was fully designed. Figure 8 shows that industries with faster employment growth rates between 1980 and 2000 have tended to locate disproportionately in metropolitan areas, which supports the nursery city view.

Another way in which the urban transmission of ideas can increase productivity is by increasing human capital acquisition for workers. This view has its roots in Marshall (1890) who wrote that in dense clusters, "the mysteries of the trade become no mystery but are, as it were, in the air." According to this view, the flow of ideas in cities enhances worker human capital. Workers learn skills directly from each other. Proximity also enables them to observe mistakes and successes and to adjust accordingly.

All of the agglomeration theories that emphasize idea transmission suggest that cities and human capital are complements, because higher levels of skills will mean more knowledge to be transmitted. In the case of innovation, more skilled workers and firms are more likely to be innovators. In the case of urban human capital accumulation, a more skilled work force means that there are more potential teachers for young workers learning from their urban peers. There is no real distinction between idea-based agglomeration economies and what non-urban scholars often call human capital externalities.

Three separate bodies of empirical evidence bear on the role of cities as disseminators of knowledge. One body of evidence looks at wage levels, particularly the connection between incomes and area-level human capital. A second examines the connection between city growth and skills. A third literature looks at the connection between city growth and other variables and attempts to sort out different ways in which localities can increase innovation.

Rauch (1993) began the modern empirical literature looking at the connection between wages and average human capital in an area. Rauch found a significant positive relationship—one extra year of area-level schooling increased the logarithm of wages by 0.05. He also found that higher levels of schooling increased housing costs. The major problem with interpreting this work is that unobserved human capital or other productivity variables may be higher in areas with higher levels of observed human capital.[15]

Acemoglu and Angrist (2001) use state-level mandatory schooling laws to test for the existence of these externalities. These rules force some cohorts in particular states to get more education than others, and Acemoglu and Angrist use this variation to look for human capital externalities.[16] They find little evidence for such externalities, which could mean that the Rauch results reflected endogeneity and omitted human capital quality. An alternative interpretation is that human capital externalities flow mostly from people at the top end of the human capital distribution rather than people at the bottom end of that distribution and the mandatory schooling laws affect the latter group. Moretti (2004a) may be the most significant recent work on this topic. He uses a number of different approaches, including the presence of federally-funded land grant colleges and other historical variables, and individual fixed effects to correct for individual-level heterogeneity. Moretti (2004b) looks at plant level production functions, and finds that productivity rises with area level human capital. This work supports Rauch's initial findings that these externalities are present.[17]

Glaeser and Mare (2001) take a slightly different approach to using wage data to understand the returns from learning in cities. They first try to rule out the possibility that the urban wage premium represents omitted human capital characteristics.[18] They then show that migrants who

[15] Rosenthal and Strange (2008b) look at the geographic incidence of human capital externalities, and agglomeration economies, and find that these are highly geographically localized, meaning that they attenuate greatly over space. In the same vein, Fu (2007) tries to measure the appropriate geographic radius for various types of spillovers within an urban area.

[16] Ciccone and Peri (2006) use similar estimates to look at the connection between changes in income and changes in schooling at the area level. They also find little evidence for human capital spillovers.

[17] Mas and Moretti (2008) provide particularly compelling evidence on peer effects within a firm where shift schedules are relatively random. Less productive workers become much more productive when they are scheduled alongside more productive peers.

[18] While Glaeser and Mare (2001) find little evidence that sorting impacts the urban wage premium in the U.S., Combes, Duranton and Gobillon (2008) find significant evidence for spatial sorting in French data.

come to cities do not receive the wage premium immediately. Instead, the age-earnings profile is steeper in big cities; i.e., migrants' wages grow faster than they would have otherwise. Furthermore, this effect is stronger in areas where a worker's industry is more heavily clustered (Freedman, 2008). One interpretation of these findings is that cities speed the acquisition of human capital.

The work on wages is paralleled by a research program that looks at the connection between initial levels of human capital in cities and population growth of those cities. Following the urban model, the value of having both approaches is that if human capital creates productivity-enhancing externalities, this should show up both in higher wages and in higher levels of population. Glaeser, Scheinkman and Shleifer (1995) find a significant connection between skills and growth across cities in post-war America. Simon and Nardinelli (2002) examine a longer time frame and find that cities with more skilled occupations in 1880 have grown more steadily since then. Simon and Nardinelli (1996) show similar results for the United Kingdom. Using a framework like that in Section II, these results were interpreted as suggesting that skills increase the rate of productivity growth at the local level.

More recently, Glaeser and Saiz (2003) and Shapiro (2006) have tried to understand why skills predict urban growth. Both papers find that the connection between area-level skills and area-level wages is rising, implying a connection between skills and productivity growth. These results support the view that successful cities thrive because of their ability to connect smart people.

The third body of research attempts to distinguish between hypotheses about how cities foment innovation. One view, associated with Marshall, Arrow and Romer, suggests that industrial concentrations of large firms will be more innovative. This view suggests localization externalities. A second view, associated with Jane Jacobs, argues that new ideas are formed by combining old ideas, and that urban diversity is the key to innovation.[19] A third view, following Porter (1990) argues that links with consumers are vital for generating new ideas. A fourth view,

[19] Berliant, Peng and Wang (2002) provide an elegant model that captures this idea.

from Chinitz (1961), argues that small firms are better for innovation, especially when they are not vertically integrated.

Regressions using city growth or the founding of new firms have typically been used to try to distinguish between these different views. For example, Glaeser et al. (1992) found evidence against concentration and for diversity. Henderson, Kuncoro and Turner (1995) find more support for the value of industrial concentration. Glaeser et al. (1992) also find that growth is faster in sectors with smaller firms, but this may reflect mean reversion of firm growth rather than new idea creation. Henderson (2003) takes a somewhat different approach and links plant-level productivity measures with area level characteristics using the Longitudinal Research Database. He finds that plants are more productive when surrounded by other similar plants. This finding is similar to the result that productivity rises with density. While it provides some evidence for agglomeration economies, it suffers from the problem that location is endogenous and omitted productivity variables should create a correlation between area-level concentration and measured productivity.

Overall, a large body of research is at least compatible with the hypothesis cities thrive because of their ability to spread knowledge. This view is also supported by the tendency of idea-oriented industries to cluster in urban centers (Glaeser and Kahn, 2001). Figure 9 shows the correlation between average human capital in an industry and the likelihood that industry will locate in an urban area. At least for now, cities appear to have a comparative advantage in more idea-intensive sectors.

The great improvements in information technology over past 30 years have led some to argue that the informational functions of physical proximity will eventually become obsolete. Gaspar and Glaeser (1998) question this view and argue that the important question becomes whether face-to-face communication and electronic communication are substitutes or complements. Theoretically, the two types of communication could certainly be complements, as people may expect to use both types of connection when forming relationships. Empirically, the situation is also murky as people seem to use the phone more when they are physically close and likely to meet. The example of Silicon Valley, which is both a famous geographic cluster and a center for information technology, cast doubts on the view that proximity and information technology are strong substitutes.

Glaeser and Ponzetto (2007) go further and argue that changes in information and transportation technologies have increased the returns to new ideas. If new ideas are best created in cities, where people can readily learn from one another, then technological changes that increase the returns to new ideas will only make cities more important, at least when those cities specialize in creating ideas. This theory can help explain why idea-oriented agglomerations, like those that specialize in financial services, have thrived over the last 30 years while goods-oriented agglomerations, like those that specialize in manufacturing, have faltered.[20]

Public Policy and Agglomeration Economies

While the literature may have converged on the view that agglomeration exist, it is not clear what, if any, policy implications come from that view. Such agglomeration economies are externalities, and externalities often imply that the unfettered market will not produce a social optimum. The presence of agglomeration economies naturally leads to the possibility of multiple equilibria and therefore suggests that small changes to initial conditions can significantly affect subsequent urban growth (Krugman, 1991a). As such, a small push could create big benefits, but what form should such small pushes take?

Some countries, but not the United States, have followed more aggressive regional policies. For example, the European Union has long directed resources towards more disadvantaged areas. In the U.S., regional policy has been limited to modest interventions in Appalachia, various attempts at urban renewal and Empowerment zones. All of these interventions were meant to strengthen troubled areas.

However, while the existence of agglomeration economies suggests that the free market may get things wrong, this doesn't imply the appropriate form of government intervention. For example,

---

[20] This model is closely tied to Duranton and Puga (2005) who show how and why cities have moved from sectoral to functional specialization.

assume that a country had two places—one poor and the other rich. The poor place is less populated, while the rich place is denser. Do agglomeration economies imply that the government should create subsidies that induce either people are firms to move from one place to another?

No. If agglomeration economies exist, then moving people out of one area will reduce the productivity of that area and increase the productivity of the receiving place. One area gets more productive and the other gets less productive. The existence of agglomeration economies does not imply that the winning area will gain more than the losing area loses. As Glaeser and Gottlieb (2008) discuss, any firm conclusions depend on the functional form of agglomeration economies. Even if we have reached a consensus that agglomeration economies exist, we have not reached any consensus over their functional form. Glaeser and Gottlieb find little clear evidence on that shape of any agglomeration effects.

The existence of human capital externalities, likewise, provides little guidance about whether more skilled workers should be pushed into already skilled areas or dispersed throughout the country. Again, the policy implication depends not only on the existence of spillovers, but on their functional form. If anything, the empirical results suggest that productivity increases with concentration of talent, but we are far from confident about that finding. If human capital spillovers are real, then this may suggest that subsidizing education might increase social welfare, but it does not suggest a regional policy that would push talent in one direction or another.

The results of this literature may, perhaps, offer more guidance to local policy-makers interested in boosting local productivity. If agglomeration economies exist, then local productivity would increase if local leaders are able to attract more economic activity. However, since attempts to attract such activity are rarely costless, the mere existence of agglomeration economies does not imply any particular policy action. They certainly do not, on their own, make the case for local subsidies to encourage new businesses. Greenstone, Hornbeck and Moretti (2007) provide the best evidence on this question.

The existence of human capital externalities does suggest that attracting skilled workers may increase local productivity and local growth. However, again, the existence of those externalities does not suggest which policies will attract skilled workers or whether such policies carry benefits that will offset their costs. Those results do suggest that there are costs associated with policies that repel highly skilled workers, such as progressive taxation at the local level.

IV. The Implications of Urban Economics for Other Fields

We now turn to the implications that urban economics has for other fields. Some of these insights reflect empirical findings in the field. Others follow from the core ingredients of urban theory.

The Economics of Growth: Human Capital and Urbanization

Much of the recent work on cities has been motivated by a desire to shed more light on the processes that also drive national economic growth. While within-country research requires somewhat different methods than cross-country methods, there are a number of findings from city-level research that seem quite relevant for cross-country researchers. The connections between area-level human capital and urban success supports the cross-county work showing a correlation between initial human capital and later GDP growth (e.g. Barro, 1991; Mankiw, Romer and Weil, 1993). There are, of course, papers (e.g., Hausmann, Pritchett and Rodrik, 2004) challenging the view that human capital increases growth rates. Others suggest that human capital is endogenous, caused by institutions and cannot be seen as a driving force of growth. The city-level evidence does not disprove these arguments, but it does support the view that education is an important determinant of area-level productivity and growth.

Indeed, a comparison of individual, metropolitan area and national income regressions shows that the correlation between human capital and incomes gets stronger at higher levels of aggregation. This is compatible with the presence of human capital spillovers that create a social multiplier, which causes group coefficients to be larger than individual level coefficients. If education improves the quality of political outcomes, then this will create another channel whereby living around skilled people can enhance productivity. Within-country evidence also supports the view that there is a link between area-level skills and reduced corruption (Glaeser and Saks, 2006).

The urban growth literature has other implications for broader growth theory. The evidence that growth at the city-industry level was negatively associated with initial scale and positively associated with small firm size casts some doubt on the view that large firms in big single industry clusters are particularly innovative. This contradicts some of the early growth models that emphasize a positive connection between innovation and either monopoly or initial scale. Following Barro and Sala-i-Martin (1991), cross-city work on income convergence can be helpful for understanding income convergence at the national level.

The urban literature also should remind us of the enormously strong connection between urbanization and income across space. Figure 10 shows the nearly-perfect connection between the logarithm of per capita GDP and urbanization rates across countries. Without sub-national evidence it would be easy to believe that this correlation was spurious, reflecting only the tendency of some countries to both be richer and have less farming. However, the within-country evidence showing a strong positive connection between density and productivity (Ciccone and Hall, 1996, Combes et al., 2008) makes that view less tenable. Of course, this does not necessarily mean that countries should subsidize urbanization, but rather that the transition to dense, urban living seems to be part of the process of countries becoming richer over time.

Idea Spillovers and Variation across Space in Innovation

The literature on knowledge spillovers suggests that ideas move quickly from person to person within urban areas. These spillovers seem to be the source of intellectual change, as urban innovators riff off each others' ideas. The role of cities in creating chains of innovations can be seen historically in events like the Florentine Renaissance, where the architect Brunelleschi developed linear perspective, which was then used in low relief sculpture by his friend Donatello, then in painting by Masaccio, and passed to his student Filippo Lippi and others. Urban intellectual connections create agglomeration economies and help us to understand why skilled cities are so successful, but they also remind us that many intellectual revolutions involve small numbers of connected inventors.

The presence of social interactions, in any context, can create high levels of random variation across space (Glaeser, Sacerdote and Scheinkman, 1996). Intuitively, when individuals act independently random variation averages out quickly—the variance of city-level averages of independent individual level outcomes is 1/N times the variance of individual level outcomes. But when outcomes are not independent, and are instead connected through urban interactions, then the variation of group-level averages can be much higher. For example, suppose the individuals in a city are connected in a circle and each person's outcome equals an independent draw plus a times the outcome of the person next to him, where $0<a<1$. With these interactions, the variance of the population average is approximately $1/(1-a)$ times 1/N time the individual variance.

The social nature of innovation can also help us think more clearly about major events in economic history, such as the industrial revolution. The industrial revolution required a number of important innovations, which were generally produced by geographically proximate inventors who borrowed each other's ideas. For example, the water frame, a machine for mechanically pulling wool, was the basis of Richard Arkwright's factory and fortune. Arkwright got the idea over drinks from a clockmaker, John Kay, who was working on a similar machine with his neighbor, the inventor Thomas Highs. Highs himself probably got the idea for the machine from earlier work done by Lewis Paul and John Wyatt. The James Watt steam engine, a similarly significant invention, came out of early collaborations in Glasgow between Watt and scientists such as John Robison and Joseph Black, and later collaborations with Matthew Boulton and William Murdoch in Birmingham.

When seminal economic innovations reflect the acts of a small number of people who learned from each other, we should expect a great deal of variation over time and space in innovative episodes like the industrial revolution or the rise of Silicon Valley. These insights are important for the vast literature that attempts to look backward and understand why events—like the Industrial Revolution—happened in places like England. If these episodes are the result of small numbers of people interacting with each other, then we would expect them to reflect random variation as much as any obvious cause. This is not to say that English institutions and canal networks were not necessary conditions for the Industrial Revolution, but it does suggest that random sparks of genius that then connected innovators across space might have been just as important.

The Spatial Equilibrium and Empirical Work on Sub-National Data

The spatial equilibrium assumption suggests that researchers cannot immediately apply tools used at the national level to sub-national data. Regressions that look at changes in income, which make perfect sense at the national level where populations are relatively fixed make far less sense in sub-national regressions where population responses are expected to be large. This does not mean that there is no role for regressions explaining income growth at the sub-national level, but they must be used together with regressions explaining population growth. Ideally, researchers would look at housing prices as well.

These facts are important for growth economists trying to use sub-national data, but they are also important for researchers in public finance who are interested in the impact of different local policies. There is a large literature that examines state-by-state variation in different policies. In some cases, the dependent variables in these regressions are very specific outcomes, like the level of private insurance. In these cases, the state policy interventions are probably too small to create a meaningful migration response. However, if a state policy is large enough to potentially impact income in a meaningful way, then it is also large enough to meaningfully change the population.

To take a concrete example, consider the pioneering work of Thomas Holmes (1997) on state right-to-work regulations. He looks at the differences in employment around state borders and connects this with different regulatory regimes in different states. He looks at employment and not incomes, which we would expect to be quite similar since workers can readily commute across borders. The work of Nathaniel Baum-Snow (2007a) and Duranton and Turner (2008) on highway production similarly looked at population rather than income effects. If a highway makes a region more productive, then we will see an increase in population and employment as long as housing supply is at least somewhat elastic—a particularly reasonable assumption in metropolitan areas recently connected to their suburbs. But, as the model shows, we will not necessarily see an increase in income.

When will a boost to local productivity have a greater impact on income as opposed to employment? This depends on the supply of net migrants, and in turn on the elasticity of housing supply since the number of people in an area is essentially proportional to the number of homes. We know little about the supply of migrants, but we are increasingly coming to understand regional differences in housing supply elasticity. The growing areas of the Sunbelt, such as Houston, Phoenix and Atlanta, appear to have abundant land and permissive land use regulations. The more static areas of the northeast and coastal California appear to have much less ability to build housing, primarily due to land use restrictions. Declining cities have a fixed supply of housing in the short run, because housing is durable (Glaeser and Gyourko, 2005).[21]

These different housing supply elasticities imply that empirical work on sub-national data may need to think seriously about how productivity increases will have different impacts in Houston and Boston. In Houston, we would expect anything that makes the region more productive to show up primarily as an increase in new construction. In Boston, where the housing supply is much more restricted, an increase in productivity should show up in higher incomes and housing prices but cannot increase population. Indeed, the economic resurgence of the Boston region since the 1970s shows up primarily in wages and prices—not population—consistent with the view that the region's housing supply is quite inelastic.

National Inequality, Housing Prices and Commute Times

The spatial equilibrium assumption and the large cross-national differences in incomes and housing prices may also be important when thinking about national income accounts and within-country inequality. If labor were perfectly mobile, and if individuals were identical, cross-city

---

[21] Helsley and Strange (1991) explain the implication of durable urban assets, like housing, for capital markets.

differences in incomes would create a false impression of income inequality with a country. Under less restrictive assumptions, the failure to think fully about space will tend to make within-country inequality estimates overstate the level of real income inequality.

For example, in 1980, the average San Francisco family was about 10 percent richer than average Houston family. In 2000, the average San Francisco family was about 50 percent richer than the average Houston family. These gaps may tell us about rising productivity in San Francisco, and about the selection of highly skilled people into that area, but they do not tell us about real income inequality. After all, the price of San Francisco housing went from being less than double the price of Houston housing to more than four times the price of Houston housing over the same time period. Using national individual data that don't correct for differences in local prices will tend to overstate real income inequality because high-wage areas also have high prices.

Following this logic, Moretti (2008) computes revised measures of the college wage premium. He finds that the high and increasing correlation between skills and house prices causes economists to significant overestimate the level and growth of U.S. inequality, but that local price indices can help correct these problems. If we use the American Chamber of Commerce Research Association metropolitan area price indices, we find that the average income in San Francisco is much closer to the average income in Houston. Moretti adjusts income with housing prices and finds that half of the increase in the college premium from 1980 to 2000 was absorbed by a higher cost of living. But even these price indices will not adjust perfectly if the unobserved amenity flows, like climate and average commute times, are different across areas.

Metropolitan area price and income differences also may cause mismeasurement of national income trends. The standard technique is to correct national average incomes with national average prices. This is not the same thing as computing average real incomes. For example, consider two regions with equal populations over two periods. In the first period both regions had average earnings of 10,000 dollars and faced a price index of one. In the second period, the first region's prices and earnings were unchanged. The second region's income had risen to 30,000 but its price index had only risen to 2, reflecting lower demand for that region's amenities. Standard accounting would see an increase in national income from 10,000 to 13,333, but average real incomes have only increased to 12,500. The problem results from applying region one's low prices to the large number of nominal dollars spent in region two. Similar problems occur if we apply national price indices to the median household if the median household lives in a relatively low-cost region of the country.

Macroeconomics, Finance and Housing Prices

Urban economics treats the price of housing as a reflection of the demand for a particular place and the costs of supplying homes. Yet this perspective rarely finds its way into the financial and macroeconomic approaches to housing prices, which tend to treat houses as they would any other asset. There is no question that macroeconomic variables influence housing prices, but the mapping is more complicated than a simple financial approach to housing would suggest. Moreover, non-financial variables, specifically the differences in productivity over space, have a large impact on aggregate housing values.

To illustrate these points, consider a simple two-region economy. In the first region, housing supply is growing and perfectly elastic. Homes are supplied by builders at a fixed construction cost of 100,000 dollars per year. In the second region, housing is completely inelastic. The price of housing will be set so that individuals are indifferent between the two locations. We assume that people are identical and risk neutral, there is a spatial equilibrium, and that everyone is a homeowner. The income difference between the two areas, x, follows a random walk.[22]

The random walk assumption implies that expected housing price appreciation will be the same in the two areas.[23] If the user cost of housing, which combines the interest rate, property taxes and maintenance costs, is 10 percent, then an income difference of x implies a housing price difference of $10x$ in the inelastic region. Thus, if income in the first region is 50,000 and income

[22] In fact, income differences across space mean revert quite strongly. This is one of the reasons why housing prices also mean revert (Glaeser and Gyourko, 2007).

[23] We are ignoring the possibility that the implied housing price in the inelastic community may become negative.

in the second region is 80,000, then house prices in the second region will be 400,000 (100,000 plus 30,000 times 10).

One implication of this type of calculation is that using the share of expenditures on housing to judge affordability of housing across space is highly misleading. In the example, housing costs are 20 percent of total income in the lower income area. Housing rises to 50 percent of total income in the higher income area. Usual affordability metrics assume that if housing costs more than thirty percent of income, then housing has become unaffordable, so the high-cost, high-income region would be classified as unaffordable but the other region would not. Yet the two areas offer exactly the same real incomes and lower housing costs in the high-income area would violate the spatial equilibrium condition.

What does this approach imply for the dynamics of national housing prices? First, average housing prices are going to be driven by the difference in income between the first and second regions. If, for example, one-half of the population lives in both regions, and the income in the first region goes down by 5,000 while the income in the high income region goes up by 5,000 dollars, then national housing prices will increase by 50,000 dollars. If incomes in both regions rise or fall by the same amount, then average housing prices will remain the same. This model's urban perspective suggests that income heterogeneity, not average income, drives national prices.

The model also offers a slightly different perspective on the impact of interest rates. If the second region's income is 5,000 dollars higher than the first region's housing prices, then a one percent decrease in interest rates, which causes the user cost of housing to decline from 0.1 to 0.09, will cause the difference in housing prices to rise by 5,556 dollars. Since the price is housing in the second region is fixed by supply, then average prices in the country will rise as interest rates fall, which is the standard comparative static illustrated by Poterba (1984) and others.

However, prices can also fall as interest rates fall. Assume that income in the second (inelastic) community is $5,000 less than in the first community, so that if the user cost is 0.1 then the price of housing is 50,000 dollars less. In this case, a one percent decrease in the interest rate that pushes the user cost down to 0.09 will cause the housing prices difference to increase by 5,556 dollars, but that can only happen if the price in the lower-income community falls by 5,556 dollars.

Other relatively counterintuitive results come from the urban economics view that emphasizes the non-financial aspects of housing. As Sinai and Souleles (2005) note, we all come into the world needing to buy housing. Owning a home, according to this perspective, is as much a hedge against future housing price shocks as a risky investment. If people are infinitely lived and never intend to move, then home ownership is a completely safe strategy while renting is risky. Housing price risk for owners becomes more severe as people expect to move between areas, change housing consumption, or perish.

This perspective also questions the view that we should expect the country in aggregate to consume significantly more as housing prices go up due to a wealth effect. An infinitely-lived consumer who wants to stay in his home has gotten neither wealthier nor poorer when housing prices rise or fall. For every home owner who sells his house in the face of a rising market, there is a buyer who is going to have to pay more for his home. Rising housing prices are better seen as a transfer from prospective buyers to perspective sellers than a nationwide increase in wealth. There are ways that rising housing prices could have a wealth effect for the aggregate economy, but they would require current owners to be more sensitive to rising prices than prospective buyers.

Understanding Local Poverty: Implications for Federalism

Economists usually see poverty as reflecting something negative about an economy that makes it unable to offer better labor market returns for poorer people. When people are mobile, however, high poverty rates in cities tell us that cities are attracting poor people—presumably by offering them something. Indeed, poverty rates in central cities are just as high among recent migrants as they are among long term residents, supporting the view that poor people are being drawn to these areas. LeRoy and Sonstelie (1979) argue that the location choices of the poor reflects high levels of demand for public transportation; Glaeser, Kahn and Rappaport (2008) provide empirical support for that claim, by showing the connection between transit access and poverty and demonstrating that increases in poverty accompany new urban transit stops.

There is also a healthy literature on welfare magnets debating whether increased transfer payments to the poor in an area increases local poverty rates (Blank, 1985, Borjas, 1999). Recent research using spatial discontinuities at state borders supports the view that poverty rates are higher when redistribution is higher (Schwuchow, 2006). A classic example of this is East St. Louis, which is one of the poorest cities in America—perhaps because it is part of the St. Louis metropolitan area but has historically received higher transfers payments because it is in Illinois.

While an increase in local poverty is usually seen as a bad thing, the urban perspective questions this interpretation. After all, public transportation has a comparative advantage in helping the poor and there is no reason to oppose poor people taking public transit and moving to areas where they have access to it. Again, this provides us with a caution against looking at an area's incomes as a measure of its success.

The mobility of the poor suggests that there may be dangers in running transfer programs at the state and local level. Large welfare differences will create concentrations of poverty, and if there are also adverse consequences for children growing up in concentrated poverty then there are reasons to question a system that will pull poorer people towards particular areas. Moreover, disproportionately taxing the wealthy in one area of the country may lead to an exodus of high income people from that area for reasons that have little to do with economic efficiency. Some urban economists have therefore argued that national redistribution is less likely to cause perverse migration responses.

Urban economics also has something to contribute to thinking about national redistribution policies. If the government pays the same amount to a welfare recipient in every location, then that sum will buy much more in a low cost area and this may induce migration to lower cost regions and also create pockets of poverty. Glaeser and Gyourko (2005) suggest that the correlation of poverty and urban decline owes much to the tendency of poorer people to move to areas where housing is cheap.

In fact, the decision of whether to index transfer payments to local price levels is a somewhat complex problem that includes a number of different considerations. Popular discussions of national transfers suggest that equity requires that these transfers be higher in high cost regions, but the spatial equilibrium concept questions that perspective; people in high cost areas are presumably already getting something in exchange for paying those higher prices (Glaeser, 1998). Moreover, the government is able to buy more for the poor by transferring more to people in low cost regions (Kaplow, 1996). The mobility response to welfare payments only further confuses the situation, and seems to suggest that the tools of urban economics can be helpful in crafting national transfer policies.

V. Directions for Future Research

The study of cities is an exciting area that can help us understand some of the most central questions in economics. The differences in productivity across place provide means of testing theories about the causes of economic output more generally. The development of technology is often a local phenomenon, so understanding why some cities are so much more innovative than others can help us to get at the very roots of technological improvement.

While the field of urban research has achieved many successes over the past four decades, we deeply hope that the best work is still ahead of us. Old questions still need more compelling answers; newer questions are still waiting to be asked. One particularly important question is how to evaluate the spatial equilibrium assumption. This essay has taken the view that this assumption is the organizing principle of urban economics. Yet there is still a lack of evidence on whether this assumption holds, especially at higher frequencies. We particularly lack evidence on whether this assumption is valid in places, like Europe, where migration flows less readily.

Understanding the sources and nature of agglomeration economies is also a particularly central question in the field. One type of approach focuses on a particular source of agglomeration benefits, such as labor market pooling, and then attempts to fully understand that source. Despite the abundance of facts about the implications of information transfer, we do not yet know exactly what types of information and ideas are transferred better within cities than across them. A second approach focuses on understanding the total functional form of agglomeration economies. The first approach offers more hope of identification. The second approach is more likely to generate parameters of policy interest. We need more research of both types.

There is a particular need for research examining the interaction between various policies and agglomeration effects. For example, transportation investments have shown a remarkable ability to make and unmake cities. Such policies are, in some cases, plausibly exogenous and provide a means for assessing the magnitude of agglomeration effects. Understanding those effects will in turn make it easier to evaluate these policies.

Since agglomeration yields benefits if there are increasing returns in firms' production functions, a better understanding of returns to scale in the service sector would help us understand why cities continue to facilitate production in a modern economy. What are the fixed costs that drive modern firms to concentrate their production geographically?

Existing research into the dynamics of regional transitions, as well as recent attempts to improve cross-regional welfare measurements, are a promising start towards a more comprehensive understanding of urban dynamics. Good measurements of migration costs would allow us to capture the short-run costs of adjusting to productivity or amenity shocks that are only incorporated into prices over time.

Finally, there has long been too much division between housing research and mainstream urban research. This division makes little sense. It is impossible to understand much about housing markets without embedding them in an urban system. In turn, the changes to the urban system work through housing markets. Limits on new construction stem the ability of higher productivity levels to create population growth. The durability of housing limits the tendency of declining productivity to decrease population in the short run. More work is needed integrating housing markets into urban research.

VI. Conclusion

More than half of humanity now lives in cities and urban agglomerations are an intrinsically important topic. Variation across these metropolises also provides a valuable means of testing hypotheses that come out of growth theory, public economics, international trade and other fields. However, proper use of city-level data requires researchers to understand something about cities. In particular, researchers looking to use regional data must keep in mind the concept of a spatial equilibrium, which plays a role in urban economics that is similar to the no-arbitrage assumption in finance. The relatively free movement of labor across cities means that urban success will show up in some combination of higher wages, higher prices and higher population levels. Housing supply elasticities, which differ across space, will determine the impact that positive shocks have on population, prices and wages.

In this essay, we reviewed the spatial equilibrium concept and showed how it could be used to understand the rise of Sunbelt cities. We then turned to a core urban topic: agglomeration economies. While there is little doubt that such economies do exist, estimating them empirically is hard because population size and productivity are determined simultaneously. Instruments that increase housing supply or consumption amenities provide the best chance of estimating the magnitude of these externalities.

We focused on three different types of agglomeration economies. There is abundant evidence that manufacturing firms choose location to reduce transport costs, but this does not seem to be an important part of urban comparative advantage today. Today, the urban role in reducing transport costs seems to be more important for service firms. Numerous researchers have argued that dense agglomerations provide labor market pooling so that workers can move from less productive to more productive firms, yet the empirical evidence supporting this claim is still modest.

The largest body of evidence supports the view that cities succeed by spurring the transfer of information. Skilled industries are more likely to locate in urban areas and skills predict urban success. Workers have steeper age-earnings profiles in cities and city-level human capital strongly predicts income. It is possible that these effects will be reduced by ongoing improvements in information technology, but that is not certain and has not happened yet.

References

Acemoglu, Daron and Joshua Angrist. 2001. "How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws." *NBER Macroeconomics Annual 2000*: 9-59.

Albion, Robert Greenhalgh. 1938. *Square-Riggers on Schedule*. Princeton: Princeton University Press.

Alonso, William. 1964. *Location and Land Use*. Cambridge, Mass.: Harvard University Press.

Audretsch, David B. and Maryann Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." American Economic Review, 86(3):630-640.

Barro, Robert J. 1991. "Economic Growth in a Cross Section of Countries." *The Quarterly Journal of Economics*, 106(2):407-443.

Barro, Robert J. and Gary S. Becker. 1989. "Fertility Choice in a Model of Economic Growth," *Econometrica*, 57(2):481-501

Barro, Robert and Xavier Sala-i-Martin. 1991. "Convergence Across States and Regions." *Brookings Papers of Economic Activity* (1):107-158.

Baum-Snow, Nathaniel. 2007a. "Did Highways Cause Suburbanization?" *Quarterly Journal of Economics* 122(2):775-805.

Baum-Snow, Nathaniel. 2007b. "Suburbanization and transportation in the monocentric model." *Journal of Urban Economics* 62: 405-442.

Becker, Gary S. and Kevin M. Murphy. 1992. "The Division of Labor, Coordination Costs, and Knowledge." The Quarterly Journal of Economics, 107(4):1137-60.

Berliant, Marcus, Peng, Shin-Kung and Ping Wang (2002) "Production Externalities and Urban Configuration," *Journal of Economic Theory* 104(2): 275-303.

Besley, Timothy, Torsten Persson and Daniel Sturm. 2005. "Political Competition and Economic Performance: Theory and Evidence from the United States." NBER working papers, no. 11484.

Black, Sandra. 1999. "Do Better Schools Matter: Parental Valuation of Elementary Schools." *Quarterly Journal of Economics* 114(2): 577-599.

Blank, Rebecca M. 1985. "The impact of state economic differentials on household welfare and labor force behavior." Journal of Public Economics, 28(1): 25-58.

Blanchard, Olivier J. and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity*. 1992(1):1-75.

Borjas, George. 1999. "Immigration and Welfare Magnets," J*ournal of Labor Economics*, 17(4):607-637.

Borts, George and Jerome Stein. 1964. *Economic Growth in a Free Market*. New York: Columbia University Press.

Brueckner, Jan. 1987. "The Structure of Urban Equilibria: A Unified Treatment of the Muth-Mills Model," in Edwin S. Mills (ed.), *Handbook of Regional and Urban Economics, Volume II*. Cambridge: Cambridge University Press, 263-289.

Caselli, Francesco and Wilbur J. Coleman. 2001. "Cross-Country Technology Diffusion: The Case of Computers." *American Economic Review*, 91(2):328-335

Paul Cheshire & Stefano Magrini, 2006. "Population growth in European cities: Weather matters -- but only nationally," Regional Studies, Taylor and Francis Journals, vol. 40(1), pages 23-37, February.

Chinitz, Benjamin. 1961. "Contrasts in Agglomeration: New York and Pittsburg." *American Economic Review* 51(2):279-289.

Ciccone, Antonio and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review*, 86(1):54-70.

Ciccone, Antonio and Giovanni Peri (2006) "Identifying Human Capital Externalities: Theory with an Application to US Cities," *Review of Economic Studies* 73(2): 381-412.

Cobb, James C. 1982. *The Selling of the South: The Southern Crusade for Industrial Development*. Baton Rouge: Louisiana State University Press.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2009. "Estimating agglomeration economies with history, geology, and worker effects," in Edward L. Glaeser (ed.), *The Economics of Agglomeration*.

Combes, Pierre-Philippe, Gilles Duranton and Laurent Gobillon, 2008. "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics* 63(2): 723-742.

Costa, Dora L. and Matthew E. Kahn. 2000. "Power Couples: Changes In The Locational Choice Of The College Educated, 1940-1990." The Quarterly Journal of Economics, 115(4):1287-1315.

Cronon, William. 1991. *Nature's metropolis: Chicago and the great west*. New York: W.W. Norton and Company.

Davis, Donald and David Weinstein. 2002. "Bones, Bombs and Break Points: The Geography of Economic Activity," *American Economic Review* 92(5):1269-1289.

Davis, Donald and David Weinstein. 2005. “Market Size, Linkages and Productivity: A Study of Japanese Regions” in Ravi Kanbur and Anthony J. Venables (eds.), *In Spatial Inequality and Develeopment*. Oxford: Oxford University Press.

Decressin, Jorg. (1993) “Regional Labor Market Evolutions in Europe,” Harvard Ph.D. Dissertation.

Diamond, Charles A. and Curtis J. Simon. 1990. “Industrial Specialization and the Returns to Labor.” *Journal of Labor Economics*, 8(1):175-201.

DiPasquale, Denise and William Wheaton. 1996. *Urban Economics and Real Estate Markets. New York*: Prentice Hall.

Duranton, Gilles and Diego Puga. 2001. “Nursery Cities: Urban diversity, process innovation, and the life cycle of products.” *American Economic Review* 91(5):1454-1477.

Duranton, Gilles and Diego Puga, 2005. “From Sectoral to Functional Urban Specialization,” *Journal of Urban Economics* 57(2): 343-370.

Duranton, Gilles and Henry Overman, 2005. “Testing for Localisation Using Micro-Geographic Data,” *Review of Economic Studies* 72(4): 1077-1106.

Duranton, Gilles, and Matthew Turner. 2008. “Urban growth and transportation.” University of Toronto Working Paper No. 305.

Eaton, Jonathan and Zvi Eckstein. 1997. “Cities and growth: Theory and evidence from France and Japan.” *Regional Science and Urban Economics* 27(4-5):443-474.

Ellison, Glenn and Edward L. Glaeser. 1997. “Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach.” *Journal of Political Economy* 105(5): 889-927.

Ellison, Glenn and Edward L. Glaeser. 1999. “The Determinants of Geographic Concentration.” *American Economic Review Papers and Proceedings* 89(2):311-316.

Ellison, Glenn, Edward L. Glaeser and William Kerr. 2007. “What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns.” NBER working papers, no. 13068.

Fallick, Bruce, Charles Fleischman and James B. Rebitzer. 2006. “Job hopping in Silicon Valley: The micro-foundations of an industry cluster.” *Review of Economics and Statistics* 88(3):472-481.

Fogel, Robert W. 1964. *Railroads and American Economic Growth: Essays in Economic History.* Baltimore: Johns Hopkins University Press.

Freedman, Matthew L. 2008. “Job hopping, earnings dynamics and industrial agglomeration in the software publishing industry.” *Journal of Urban Economics* 64(3):590-600.

Fu, Shihe. 2007. "Smart café cities: Testing human capital externalities in the Boston metropolitan area." *Journal of Urban Economics* 61(1):86-111.

Fuchs, Victor. 1957. *The Economics of the Fur Industry*. New York: Columbia University Press.

Fujita, Masahisa, Paul Krugman and Anthony J. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Economy*. Cambridge, Mass.: MIT Press.

Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *The Quarterly Journal of Economics*, 114(3):739-767.

Gaspar, Jess and Edward L. Glaeser. 1998. "Information Technology and the Future of Cities." *Journal of Urban Economics* 43(1):136-156.

Glaeser, Edward L. 1989. "Should Transfer Payments be Indexed to Local Price Levels?" *Regional Science and Urban Economics* 28:1-20.

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

Glaeser, Edward L. and Joshua D. Gottlieb. 2006. "Urban Resurgence and the Consumer City." *Urban Studies*, 43(8):1275-1299.

Glaeser, Edward L. and Joshua D. Gottlieb. 2008. "The Economics of Place-Making Policies." *Brookings Papers on Economic Activity*, 2008(1):155-239.

Glaeser, Edward L. and Joseph Gyourko. 2005 "Urban Decline and Durable Housing." *Journal of Political Economy,* 113(2):345-375.

Glaeser, Edward L. and Joseph Gyourko. 2007. "Housing Dynamics." Harvard University, mimeo.

Glaeser, Edward L., Joseph Gyourko and Raven E. Saks. 2005. "Why Have Housing Prices Gone Up?" *American Economic Review Papers and Proceedings*, 95(2):329-333.

Glaeser, Edward L. and Matthew E. Kahn. 2001. "Decentralized Employment and the Transformation of the American City." *Brookings/Wharton Papers on Urban Affairs* 2.

Glaeser, Edward L., Matthew E. Kahn and Jordan Rappaport. 2008. "Why do the Poor Live in Cities?" *Journal of Urban Economics* 63(1) (2008):1-24.

Glaeser, Edward. L., Heidi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer. 1992. "Growth in Cities," Journal of Political Economy, 100(6):1126-52.

Glaeser, Edward L. and Janet Kohlhase. 2004. "Cities, Regions and the Decline of Transport Costs." *Papers in Regional Science* 83(1):197-228.

Glaeser, Edward L. and David Mare. 2001 "Cities and Skills." *Journal of Labor Economics* 19(2):316-342.

Glaeser, Edward L. and Charles Redlick. 2008. "Social Capital and Urban Growth." Harvard University, mimeo.

Glaeser, Edward L., Bruce Sacerdote, and Jose A. Scheinkman. 1996. "Crime and Social Interactions," *The Quarterly Journal of Economics* 111(2):507-548.

Glaeser, Edward L. and Raven E. Saks. 2006. "Corruption in America." *Journal of Public Economics* 90(6-7):959-1364.

Glaeser, Edward L. and Giacomo A.M. Ponzetto. 2007. "Did the Death of Distance Hurt Detroit and Help New York?" Harvard University, mimeo.

Glaeser, Edward L., Jose A. Scheinkman and Andrei Shleifer. 1995. "Economic Growth in a Cross-Section of Cities." *Journal of Monetary Economics* 36(1):117-143.

Glaeser, Edward L. and Kristina Tobio. 2007. "The Rise of the Sunbelt." Harvard University, mimeo.

Greenstone, Michael, Richard Hornbeck and Enrico Moretti (2007) "Bidding for Industrial Plants: Does Winning a 'Million Dollar' Plant Increase Welfare?" MIT, mimeo.

Gyourko, Joseph and Albert Saiz. 2006. "Construction Costs and the Supply of Housing Structure." Journal of Regional Science, 46(4):661-680.

Hanson, Gordon H. 2005. "Market Potential, Increasing Returns and Geographic Concentration." *Journal of International Economics*, 67(1):1-24.

Harris John R. and Michael P. Todaro. 1970. "Migration, Unemployment and Development: A Two-Sector Analysis." *American Economic Review*, 60(1):126-42.

Hausmann, Ricardo, Lant Pritchett and Dani Rodrik. 2005. "Growth Accelerations." *Journal of Economic Growth* 10(4):303-329.

Head, Keith and Thierry Mayer. 2004. "Market Potential and the Location of Japanese Investment in the European Union." *Review of Economics and Statistics*, 86(4):959-972.

Helsley, Robert W. and William C. Strange. 1990. "Matching and Agglomeration Economics in a System of Cities." *Regional Science and Urban Economics*, 20(2):189-212.

Helsley, Robert W. and William C. Strange. 1991. “Agglomeration economies and urban capital markets,” *Journal of Urban Economics*, 29(1):96-112.

Helsley, Robert W. and William C. Strange, 2004. “Knowledge Barter in Cities,” *Journal of Urban Economics* 56(2): 327-345.

Henderson, J. Vernon. 1988. *Urban Development: Theory, Fact, and Illusion*. Oxford: Oxford University Press.

Henderson, J. Vernon. 2003. “Marshall’s Scale Economies.” *Journal of Urban Economics*, 53(1):1-28.

Henderson, J. Vernon, Ari Kuncoro and Matt Turner. 1995. “Industrial Development in Cities.” *Journal of Political Economy*, 103(5):1067-1090.

Henderson, J. Vernon and Yukako Ono. 2008. “Where do manufacturing firms locate their headquarters?” *Journal of Urban Economics* 63(2):431-450.

Hoover, Edgar M. 1948. *The Location of Economic Activity*. New York: McGraw Hill.

Hornbeck, Richard. 2008. “Quantifying Long-term Adjustment to Environmental Change: Evidence from the American Dust Bowl.” MIT, mimeo.

Jaffe, Adam B., Manuel Trajtenberg and Rebecca Henderson. 1993. “Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations.” *The Quarterly Journal of Economics* 108(3):577-598.

Kaplow, Louis. 1996. “ Regional Cost-of-Living Adjustments in Tax/Transfer Schemes.” Tax Law Review. 51(2):175-198.

Kolko, Jed. 2000. *Essays on Information Technology, Cities, and Location Choice*. Harvard University, dissertation.

Krugman, Paul. 1991a. “Increasing Returns and Economic Geography.” *Journal of Political Economy*, 99(3):483-499.

Krugman, Paul. 1991b. *Geography and Trade*. Cambridge: MIT Press.

LeRoy, Stephen and Jon Sonstelie. 1983. “Paradise Lost and Regained: Transportation Innovation, Income and Residential Location.” *Journal of Urban Economics* 13:67-89.

Lucas, Robert. 1988. “On the Mechanics of Economic Development.” *Journal of Monetary Economics*, 22(1):3-42.

Mankiw, Gregory, David Romer and David Weil. 1992. “A Contribution to the Empirics of Economic Growth.” *The Quarterly Journal of Economics*, 107(2):407-437.

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

Mas, Alexandre and Enrico Moretti (2008) "Peers at Work," *American Economic Review*, forthcoming.

Mills, Edwin S. 1967. "An Aggregative Model of Resource Allocation in a Metropolitan Area." *American Economic Review*, 57(2):197-210.

Moretti, Enrico. 2004a. "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics* 121(1-2):175-212.

Moretti, Enrico. 2004b. "Workers' Education, Productivity and Spillovers: Evidence from Plant-Level Production Functions," *American Economic Review* 84(3): 656-690.

Moretti, Enrico. 2008. "Real Wage Inequality." University of California, Berkeley, mimeo.

Mueser, Peter and Philip Graves. 1995. "Examining the Role of Economic Opportunity and Amenities in Explaining Population Redistribution." *Journal of Urban Economics* 37(2):176-200.

Muth, Richard. 1969. *Cities and Housing*. Chicago: University of Chicago Press.

Olson, Mancur. 1983. "The South Will Fall Again: The South as Leader and Laggard in Economic Growth." *Southern Economic Journal* 49(4):917-932.

Overman, Henry and Diego Puga. 2007. "Labor Market Pooling as a Source of Agglomeration: An Empirical Investigation." London School of Economics, mimeo.

Porter, Michael E. 1990. *The Competitive Advantage of Nations.* New York: Free Press.

Poterba, James. 1984. "Tax Subsidies to Owner-Occupied Housing: An Asset Market Approach," *The Quarterly Journal of Economics* 99(4):729-745.

Rappaport, Jordan and Jeffrey Sachs. 2003. "The United States as a Coastal Nation." *Journal of Economic Growth*, 8(1):5-46.

Rauch, James E. 1993. "Does History Matter Only When It Matters Little? The Case of City-Industry Location." *The Quarterly Journal of Economics* 108(3):843-867.

Roback, Jennifer. 1982. "Wages, Rents and the Quality of Life." *Journal of Political Economy* 90(6):1257-1278.

Romer, Paul M. 1986. "Increasing Returns and Long-run Growth." *Journal of Political Economy*, 94(5):1002-1037.

Rosen, Sherwin. 1979. “Wage-based indexes of urban quality of life.” In Peter Miezkowski and Mahlon R. Straszheim, eds., *Current Issues in Urban Economics*. Baltimore: Johns Hopkins University Press.

Rosenthal, Stuart and William C. Strange. 2008a. “Agglomeration and Hours Worked,” *Review of Economics and Statistics*, 90:105-115.

Rosenthal, Stuart and William C. Strange. 2008b. “The Attenuation of Human Capital Spillovers. *Journal of Urban Economics*, 64:373-389.

Rosenthal, Stuart S. and William C. Strange. 2003. “Geography, Industrial Organization, and Agglomeration.” *The Review of Economics and Statistics*. 85(2):377-393.

Saks, Raven E. and Abigail Wozniak. 2007. “Labor Mobility Over the Business Cycle: New Evidence from Internal Migration.” University of Notre Dame, mimeo.

Saxenian, AnnaLee. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, Mass.: Harvard University Press.

Schwartz, Amy Ellen, Scott Susin and Ioan Voicu. 2003. “Has Falling Crime Driven New York City’s Real Estate Boom?” *Journal of Housing Research***,** 14(1):101-135.

Schwuchow, Timothy J. 2006. “Do Differences in State Welfare Policy Affect the Residential Choices of the Poor?” Harvard University, thesis.

Shapiro, Jesse M. 2006. “Smart Cities: Quality of Life, Productivity, and the Growth Effects of Human Capital.” *Review of Economics and Statistics* 88(2):324-335.

Simon, Curtis J. and Clark Nardinelli. 1996. “The Talk of the Town: Human Capital, Information, and the Growth of English Cities, 1861 to 1961.” *Explorations in Economic History* 33(3):384-413.

Simon, Curtis J. and Clark Nardinelli. 2002. “Human Capital and the Rise of American Cities,” Regional Science and Urban Economics xxx.

Sinai, Todd and Nicholas Souleles. 2005. “Owner-Occupied Housing as a Hedge Against Rent Risk.” *Quarterly Journal of Economics*, 120(2):763-789.

Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan and T. Cadell.

Solow, Robert M. 1956. “A Contribution to the Theory of Economic Growth.” *The Quarterly Journal of Economics*, 70(1):65-94.

Strange, William C., Hejazi, Walid and Jianmin Tang. 2006. “The Uncertain City: Competitive Instability, Skills, Innovation and the Strategy of Agglomeration,” *Journal of Urban Economics* 59: 331-351.

Thaler, Richard H.. 1978. "A Note on the Value of Crime Control: Evidence from the Property Market." *Journal of Urban Economics*, 5: 137-145.

von Thunen, Johan. 1826. *The Isolated State*. Hamburg: F. Perthes.

Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort*. Cambridge, Mass.: Addison Wesley Press.

Appendix: Data Description

Census aggregated data are taken from the compilation provided by the Inter-university Consortium for Political and Social Research, under record number 2896. This compilation includes Census county data from 1790 to 2000, including data from the Census's various City and County Data Books.

To analyze metropolitan-area level data, we aggregate the county data according to Metropolitan Statistical Area (MSA) definitions released by the Office of Management and Budget. Each figure or table specifies the definition used for that particular application. We use different definitions for different purposes in order to be consistent with data from other sources used in a particular figure or table.

A word of caution is in order regarding some aggregate numbers computed from these data. In order to use a consistent set of MSA definitions for each purpose we need median family income and median house value data at various MSA definitions, these medians are only presented under certain definitions. We therefore estimate the median by averaging the component counties' median values, weighting by families in the case of family income and by housing units in the case of house values. The resulting numbers are not equal to the true median for the metropolitan area, but they should be a close enough approximation for our purposes.

We obtain the Census Bureau's 5% Public Use Microdata Sample (PUMS) of the 2000 Census from the Integrated Public Use Microdata Series (IPUMS) service of the Minnesota Population Center. The sole geographical identifier included in the PUMS is a Public Use Microdata Area (PUMA), which IPUMS links to an MSA where appropriate. (In particular, IPUMS uses the 1999 MSA definitions, using Primary Metropolitan Statistical Areas rather than Consolidated Metropolitan Statistical areas where applicable.) This identification is imperfect because the Census does not ensure that each PUMA is contained within a county, so PUMAs do not necessarily map to MSAs. Nonetheless it is the best that can be done to link Census microdata to other geographical data.

When we use industry-level data in conjunction with Census industry categorization, it is necessary to match the different industry classification systems used in the different datasets.

Census industry codes for manufacturing industries, on the 1990 basis, are matched to Standard Industrial Classification (SIC) codes using Appendix A to Census Technical Paper No. 65, which is available online at http://www.census.gov/hhes/www/ioindex/tp65_report.html. Since there is not a one-to-one relationship between Census industry codes and SIC codes, the concordance is necessarily imperfect; we select one SIC code if there are multiple ones given, and we use data from the SIC code given even when informed that the Census industry code only matches part of the SIC category. A given Census industry code can be matched with a 2-digit, 3-digit or 4-digit SIC code, so our resulting dataset uses a mixture of levels of detail.

These SIC data are in turn matched to transportation cost data from the Commodity Flow Survey, which reports product-level data using the Standard Classification of Transported Goods (SCTG). We match each industry with its respective product by hand, and note that the inconsistent level of detail in our SIC data propagates into the SCTG concordance, where a given SIC code can be matched to a 2-, 3-, or 4-digit SCTG code. Furthermore, a detailed SIC code may be matched to a detailed SCTG code while a more general related SIC code may be matched to a coarser SCTG code that includes the detailed code used elsewhere.

Table 1: Convergence and growth

Decade	Income growth	Population growth
1960s	-0.26 [0.02]	0.14 [0.04]
1970s	-0.21 [0.02]	-0.18 [0.05]
1980s	-0.006 [0.035]	0.002 [0.051]
1990s	-0.090 [0.016]	-0.001 [0.034]

Note: Each cell is a separate regression of the variable listed at the top of the column on initial log income per capita. All regressions include a constant. Data are from the Census, as described in the Data Appendix. Units of observation are Metropolitan Statistical Areas under the 1999 definitions, using Primary Metropolitan Statistical Areas rather than Consolidated Metropolitan Statistical Areas where applicable and New England County Metropolitan Areas where applicable.

Table 2: Estimating Parameters		
Equation Parameters	Value of Parameters in the Baseline Model	Value of Parameters with Agglomeration Economies
λ_A^N	$\frac{\beta+\mu(1-\beta)(1-\eta)}{\big((1-\alpha)\eta+\alpha\gamma\big)\mu(1-\beta)+\alpha\beta\gamma.}$	$\frac{\beta+\mu(1-\beta)(1-\lambda)}{(1-\alpha+\omega)(1-\beta)\eta\mu+(\mu+\beta-\mu\beta)(\alpha\gamma-\omega)}$
λ_H^N	$\frac{(1-\alpha+\alpha\gamma)(1-\beta)}{\big((1-\alpha)\eta+\alpha\gamma\big)\mu(1-\beta)+\alpha\beta\gamma.}$	$\frac{(1-\alpha+\alpha\gamma)(1-\beta)}{(1-\alpha+\omega)(1-\beta)\eta\mu+(\mu+\beta-\mu\beta)(\alpha\gamma-\omega)}$
λ_θ^N	$\frac{1-\alpha+\alpha\gamma}{\big((1-\alpha)\eta+\alpha\gamma\big)\mu(1-\beta)+\alpha\beta\gamma.}$	$\frac{1-\alpha+\alpha\gamma}{(1-\alpha+\omega)(1-\beta)\eta\mu+(\mu+\beta-\mu\beta)(\alpha\gamma-\omega)}$
λ_A^W	$\frac{\mu\eta(1-\beta)}{\big((1-\alpha)\eta+\alpha\gamma\big)\mu(1-\beta)+\alpha\beta\gamma.}$	$\frac{\mu\eta(1-\beta)}{(1-\alpha+\omega)(1-\beta)\eta\mu+(\mu+\beta-\mu\beta)(\alpha\gamma-\omega)}$
λ_H^W	$\frac{-\alpha\gamma(1-\beta)}{\big((1-\alpha)\eta+\alpha\gamma\big)\mu(1-\beta)+\alpha\beta\gamma.}$	$\frac{-(1-\beta)(\alpha\gamma-\omega)}{(1-\alpha+\omega)(1-\beta)\eta\mu+(\mu+\beta-\mu\beta)(\alpha\gamma-\omega)}$
λ_θ^W	$\frac{-\alpha\gamma}{\big((1-\alpha)\eta+\alpha\gamma\big)\mu(1-\beta)+\alpha\beta\gamma.}$	$\frac{-(\alpha\gamma-\omega)}{(1-\alpha+\omega)(1-\beta)\eta\mu+(\mu+\beta-\mu\beta)(\alpha\gamma-\omega)}$
λ_A^P	$\frac{\mu\eta}{\big((1-\alpha)\eta+\alpha\gamma\big)\mu(1-\beta)+\alpha\beta\gamma.}$	$\frac{\mu\eta}{(1-\alpha+\omega)(1-\beta)\eta\mu+(\mu+\beta-\mu\beta)(\alpha\gamma-\omega)}$
λ_H^P	$\frac{\alpha\gamma}{\big((1-\alpha)\eta+\alpha\gamma\big)\mu(1-\beta)+\alpha\beta\gamma.}$	$\frac{-(\alpha\gamma-\omega)}{(1-\alpha+\omega)(1-\beta)\eta\mu+(\mu+\beta-\mu\beta)(\alpha\gamma-\omega)}$
λ_θ^P	$\frac{(1-\alpha)\mu\eta-(1-\mu)\alpha\gamma}{\big((1-\alpha)\eta+\alpha\gamma\big)\mu(1-\beta)+\alpha\beta\gamma.}$	$\frac{(1-\alpha+\omega)\eta\mu-(1-\mu)(\alpha\gamma-\omega)}{(1-\alpha+\omega)(1-\beta)\eta\mu+(\mu+\beta-\mu\beta)(\alpha\gamma-\omega)}$

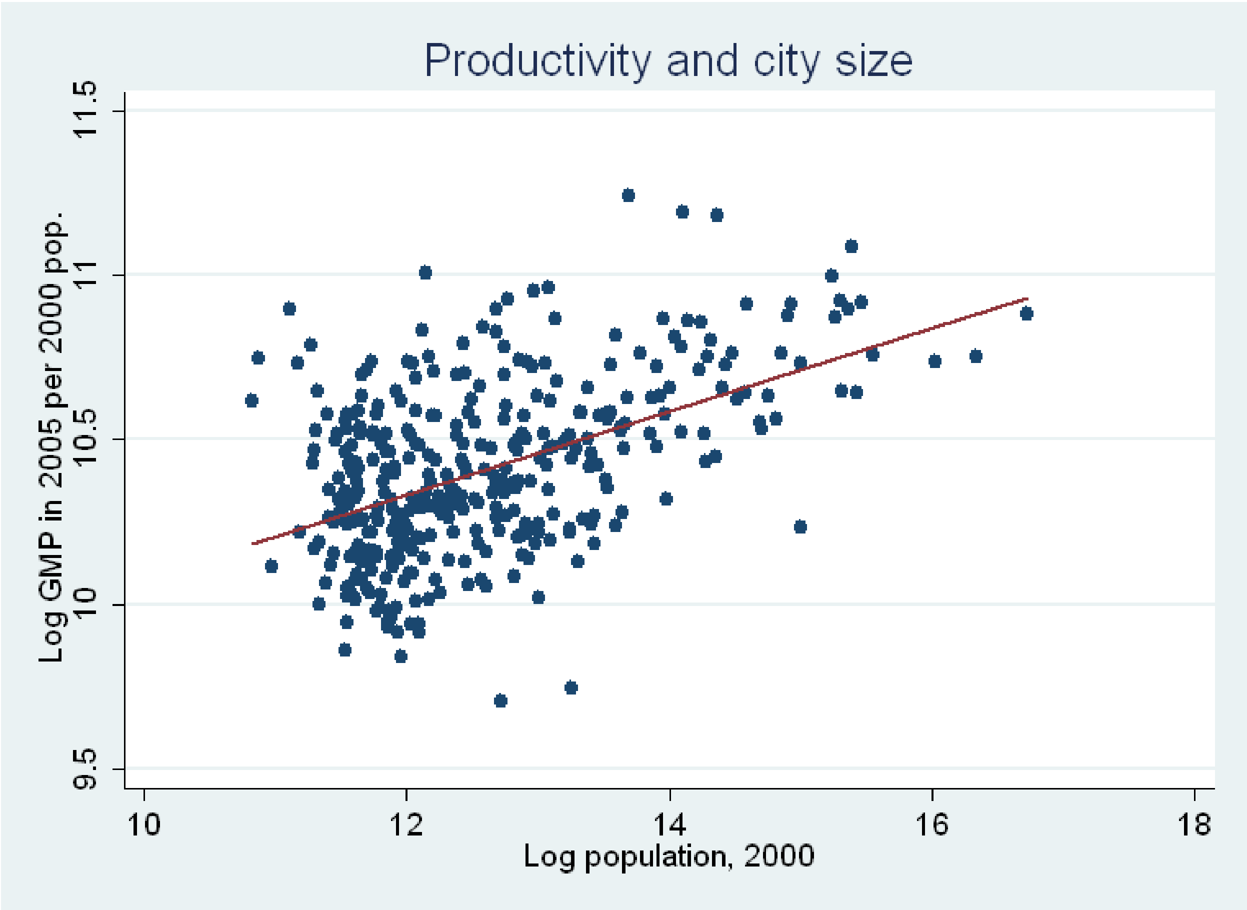
Table 3: Spatial equilibrium						
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	Log wage	Log house value	Log real wage	Log wage	Log house value	Log real wage
Year:	2000	2000	2000	1990, 2000	1990, 2000	1990, 2000
Mean January	-0.19	0.60	-0.33			
	[0.06]	[0.31]	[0.10]			
Mean January				-0.001	-0.43	0.19
				[0.05]	[0.11]	[0.03]
Year 2000 dummy				0.25	0.62	0.06
				[0.02]	[0.06]	[0.02]
Individual controls	Yes	-	Yes	Yes	-	Yes
Housing controls	-	Yes	-	-	Yes	-
MSA Fixed Effects	-	-	-	Yes	Yes	Yes
N	1,590,467	2,341,976	1,590,467	2,950,850	4,245,315	2,950,850
R^2	0.29	0.36	0.21	0.27	0.60	0.26

Note: Individual-level data are from the Census Public Use Microdata Sample, as described in the Data Appendix. Metropolitan-area population is from the Census, as also described in the Data Appendix. Mean January temperature is from the City and County Data Book, 1994, and is measured in hundreds of degrees Fahrenheit. Real wage is controlled for with median house value, also from the Census as described in the Data Appendix. Individual controls include sex, age, and education. Location characteristics follow Metropolitan Statistical Areas under the 1999 definitions, using Primary Metropolitan Statistical Areas rather than Consolidated Metropolitan Statistical Areas where applicable and New England County Metropolitan Areas where applicable. Standard errors are clustered by metropolitan area.

Table 4: City size									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable	Log wage	Log wage	Log wage	Log house price	Log house price	Log house price	Log real wage	Log real wage	Log real wage
Regression type	OLS	IV population	IV geography	OLS	IV population	IV geography	OLS	IV population	IV geography
Log population,	0.04	0.08	0.04	0.16	0.06	0.39	-0.024	0.025	-0.09
	[0.01]	[0.03]	[0.02]	[0.03]	[0.06]	[0.09]	[0.019]	[0.054]	[0.03]
N	1,591,140	1,521,599	1,590,467	2,343,054	2,220,249	2,333,002	1,591,140	1,521,599	1,590,467
R^2	0.22			0.40			0.20		

Note: Individual-level data are from the Census Public Use Microdata Sample, as described in the Data Appendix. Metropolitan-area population is from the Census, as also described in the Data Appendix. Mean January temperature, which is measured in hundreds of degrees Fahrenheit, and precipitation are from the City and County Data Book, 1994, and Fahrenheit. Real wage is controlled for with median house value, also from the Census as described in the Data Appendix. Individual controls include sex, age, and education. Location characteristics follow Metropolitan Statistical Areas under the 1999 definitions, using Primary Metropolitan Statistical Areas rather than Consolidated Metropolitan Statistical Areas where applicable and New England County Metropolitan Areas where applicable. Standard errors are clustered by metropolitan area.

Figure 1

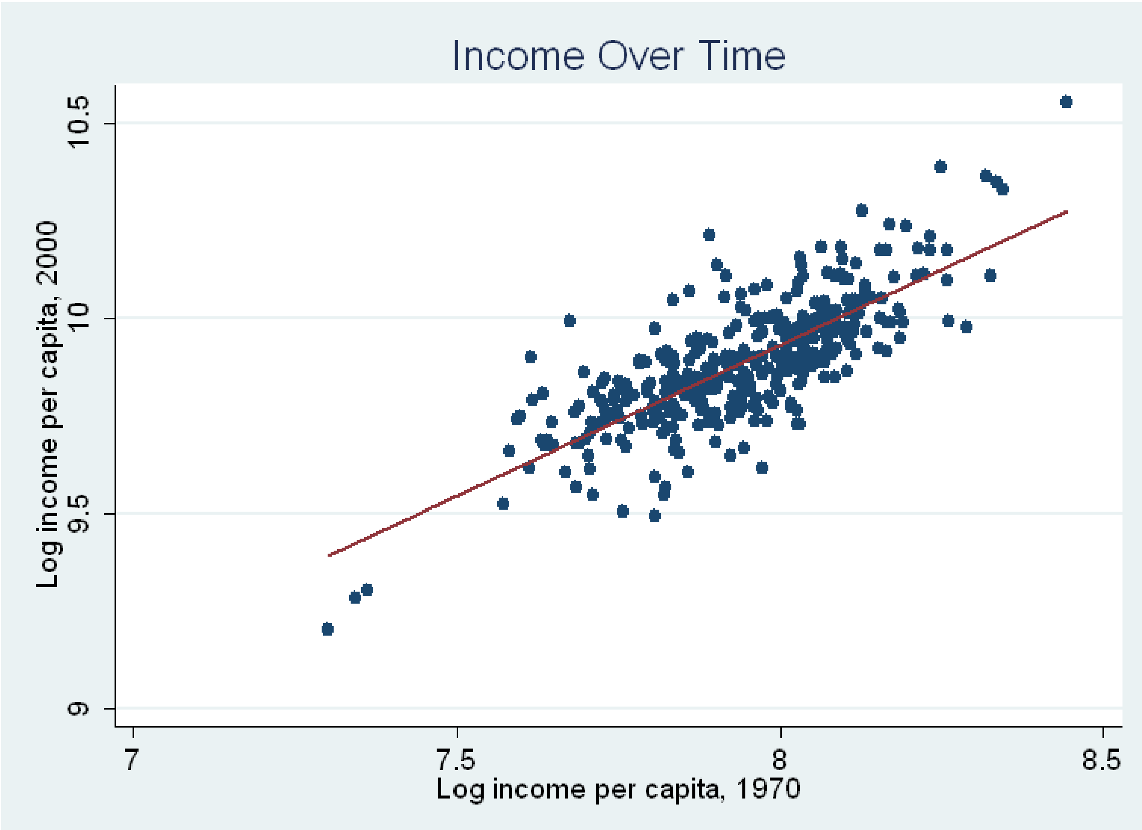


Note: Units of observation are Metropolitan Statistical Areas under the 2006 definitions. Population is from the Census, as described in the Data Appendix. Gross Metropolitan Product is from the Bureau of Economic Analysis.

The regression line is *Log GMP per capita* = 0.13 [0.01] * *Log population* + 8.8 [0.1].

R^2 = 0.25 and N = 363.

Figure 2



Note: Units of observation are Metropolitan Statistical Areas under the 2006 definitions, using Metropolitan Divisions where applicable. Data are from the Census, as described in the Data Appendix.

The regression line is *Income 2000* = 0.77 [0.03] * *Income 1970* + 3.75 [0.26].

R^2 = 0.60 and *N* = 363.

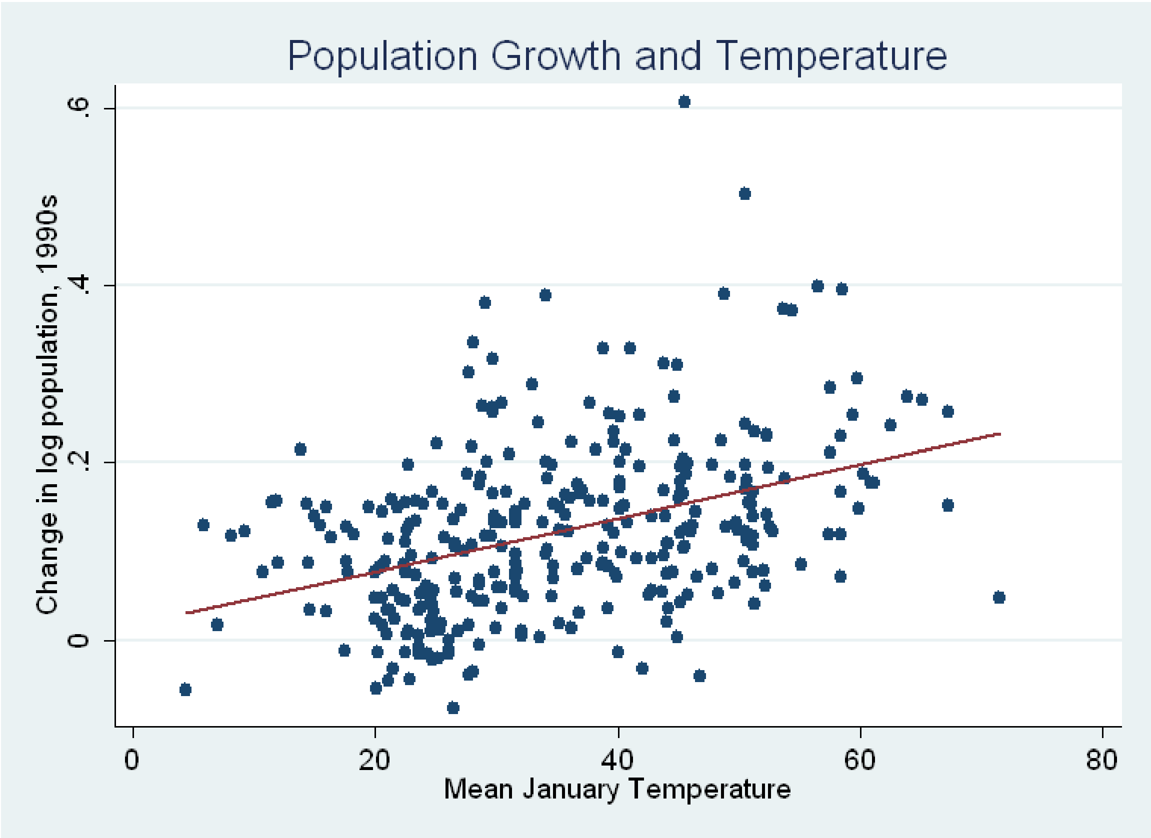
Figure 3

Note: Units of observation are Metropolitan Statistical Areas under the 2006 definitions. Data are from the Census, as described in the Data Appendix.

The regression line is *Log income* = 0.34 [0.02] * *Log value* + 5.97 [0.22].

R^2 = 0.46 and N = 363.

Figure 4

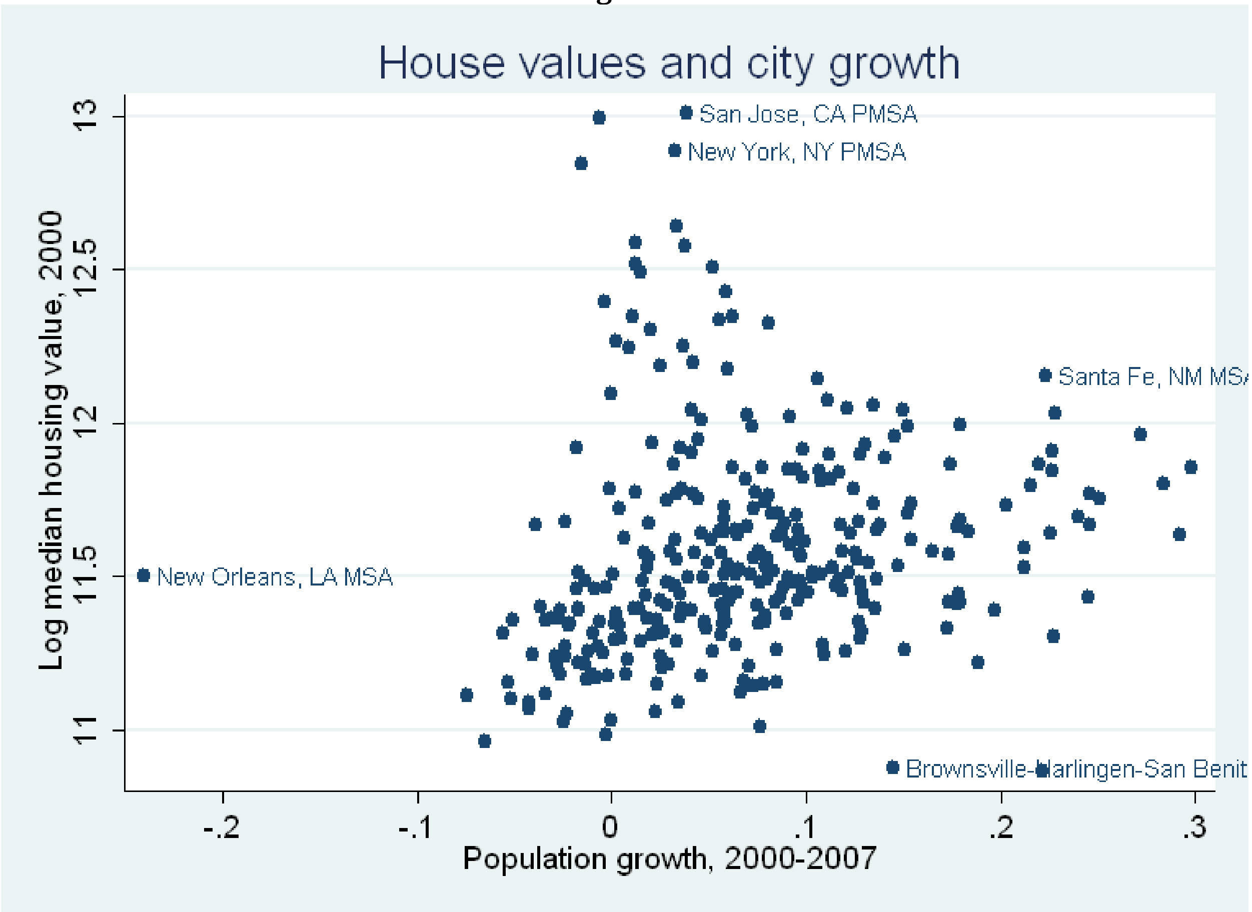


Note: Units of observation are Metropolitan Statistical Areas under the 1999 definitions, using Primary Metropolitan Statistical Areas rather than Consolidated Metropolitan Statistical Areas where applicable and New England County Metropolitan Areas where applicable. Population data are from the Census, as described in the Data Appendix. Mean January temperature is from the City and County Data Book, 1994.

The regression line is *Population growth* = 0.0030 [0.0004] * *Temperature* + 0.02 [0.01].

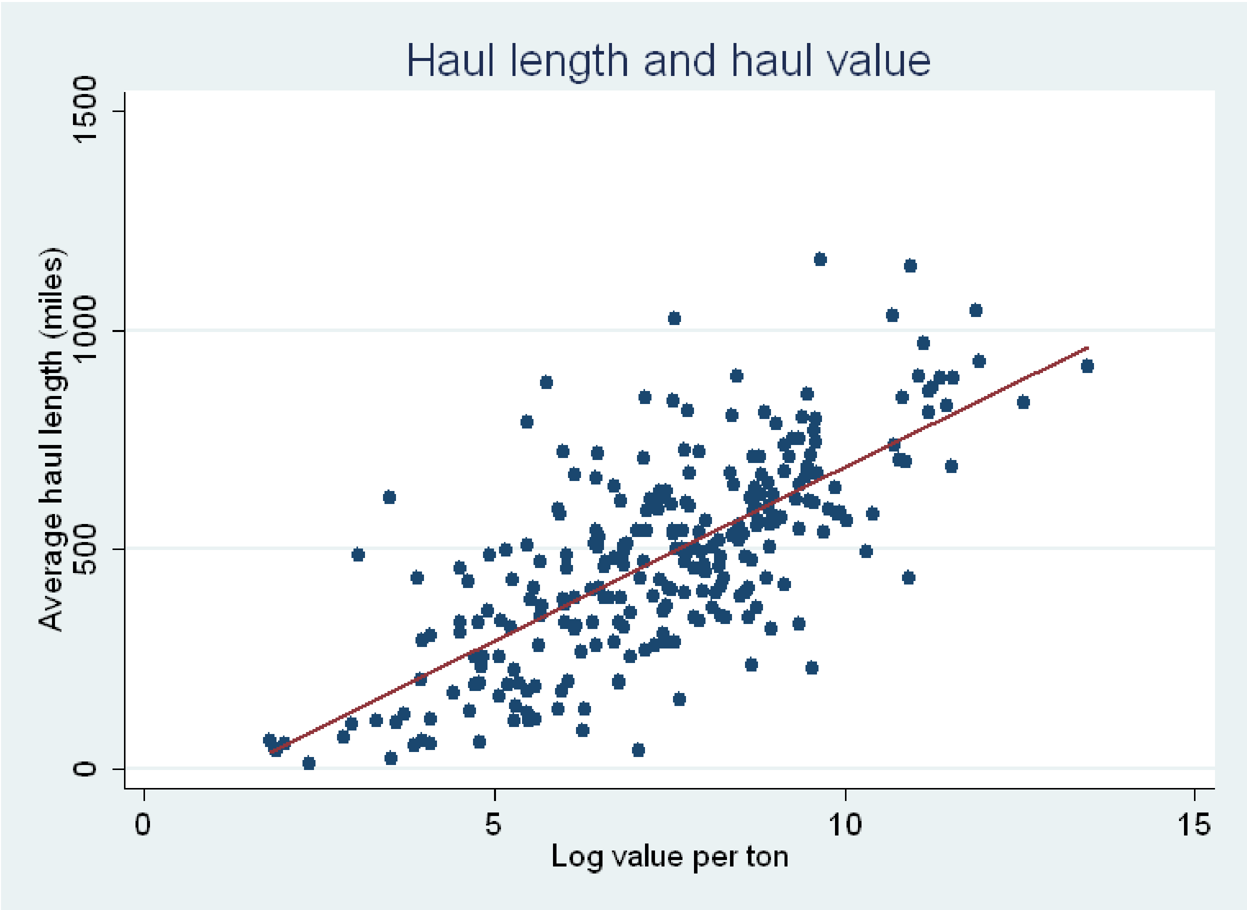
R^2 = 0.16 and N = 316.

Figure 5



Note: Units of observation are Metropolitan Statistical Areas under the 1999 definitions, using Primary Metropolitan Statistical Areas rather than Consolidated Metropolitan Statistical Areas where applicable and New England County Metropolitan Areas where applicable. Data are from the Census, as described in the Data Appendix.

Figure 6

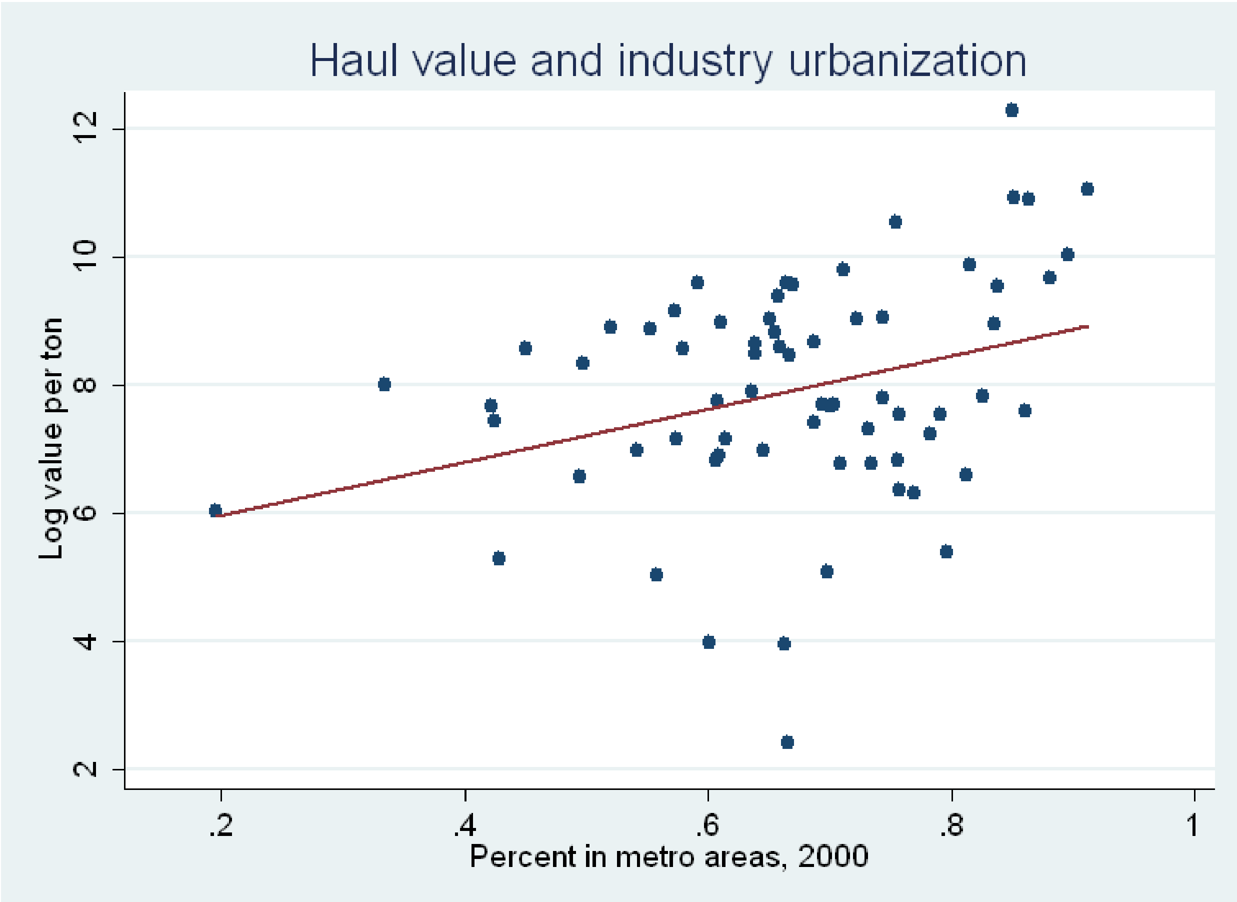


Note: Units of observation are commodities at the 4-digit Standard Classification of Transported Goods level. Data are from the 1997 Commodity Flow Survey.

The regression line is *Average length* = 79 [5] * *Log value per ton* – 107 [35].

$R^2 = 0.53$ and $N = 267$

Figure 7

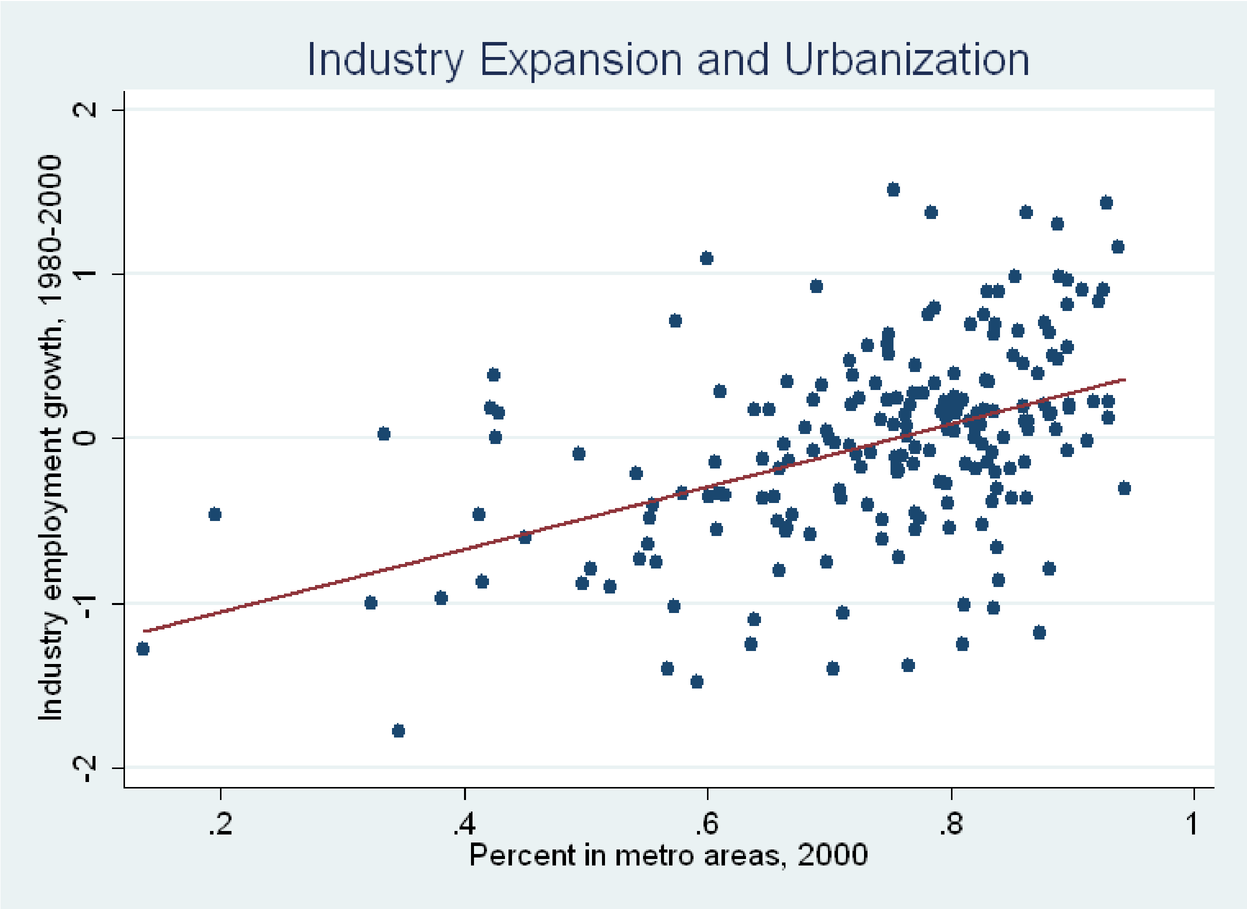


Note: Units of observation are Census industries, matched to commodities as described in the Data Appendix. Industry-level urbanization is from the Census Public Use Microdata Sample, as described in the Data Appendix. Value per ton is from the 1997 Commodity Flow Survey.

The regression line is *Log value per ton* = 4.1 [1.5] * *Urbanization* + 5.1 [1.0].

$R^2 = 0.11$ and $N = 68$

Figure 8

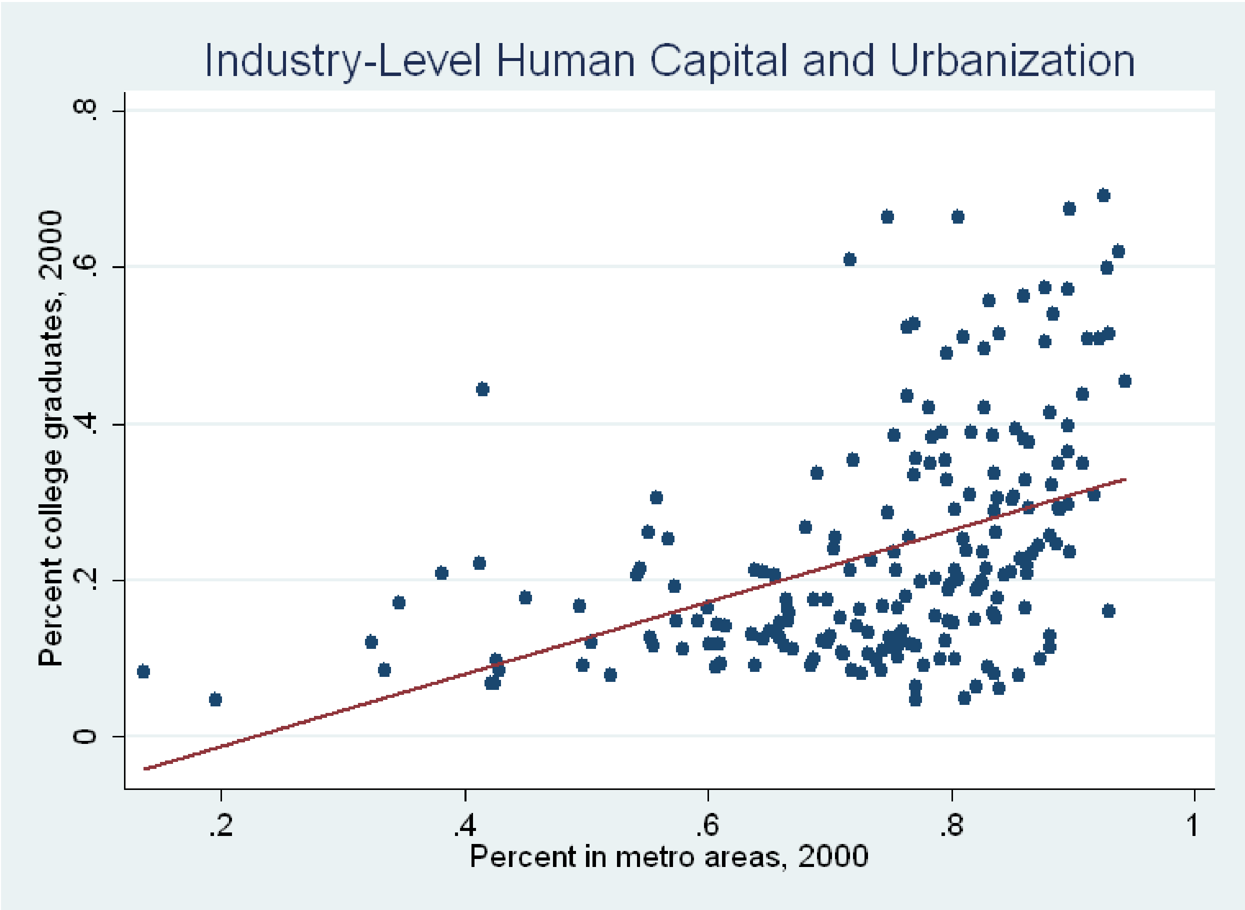


Note: Units of observation are Census industries, as described in the Data Appendix. Data are from the Census Public Use Microdata Sample, as also described in the Data Appendix.

The regression line is *Industry growth* = 1.9 [0.3] * *Urbanization* -1.4 [0.2].

R^2 = 0.21 and N = 197

Figure 9

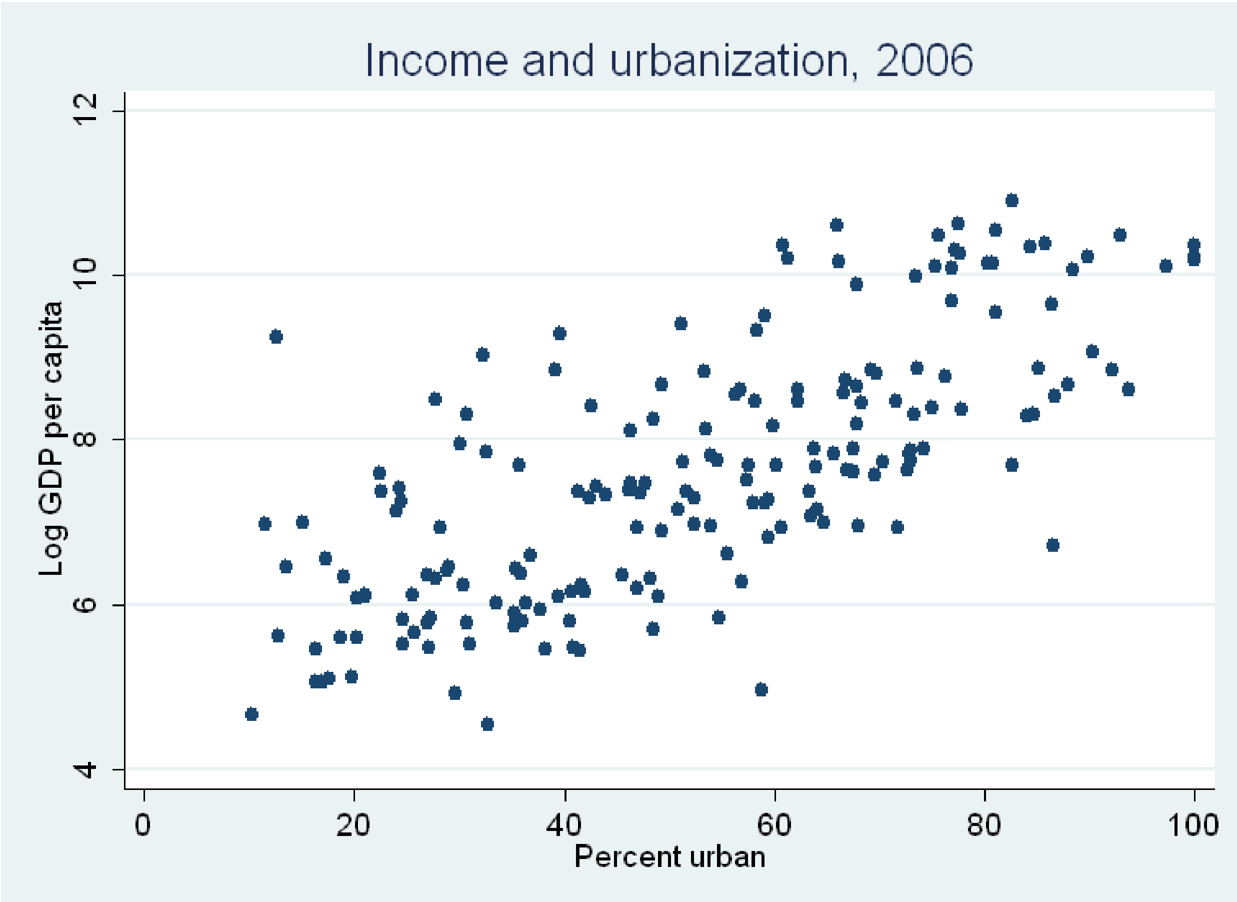


Note: Units of observation are Census industries, as described in the Data Appendix. Data are from the Census Public Use Microdata Sample, as also described in the Data Appendix.

The regression line is *Percent college graduates* = 0.46 [0.07] * *Urbanization* – 0.10 [0.05].

R^2 = 0.19 and N = 205

Figure 10



Note: Units of observation are countries. Data are from the World Development Indicators database. GDP per capita is measured in constant 2000 U.S. dollars.