Smart Cities: Explaining the Relationship between City Growth and Human Capital

Jesse M. Shapiro[*,†]
Harvard University

June 19, 2003

Abstract

From 1940 to 1990, a 10 percent increase in a metrpolitan area's concentration of college-educated residents was associated with a .6 percent increase in subsequent employment growth. Using data on growth in wages and house values, I attempt to distinguish between explanations for this correlation based on local productivity growth, and explanations based on growth in local consumption amenities. Calibration of a city growth model suggests that roughly two-thirds of the growth effect of human capital is due to enhanced productivity growth, the rest being caused by growth in the quality of life. This contrasts with the standard argument that human capital generates growth in urban areas solely through local knowledge spillovers.

From 1940 to 1990, a 10 percent increase in a metropolitan area's concentration of human capital was associated with roughly a .6 percent increase in the area's employment growth. A substantial body of literature confirms this correlation between human capital and local area growth.[1] Little is known, however, about the underlying cause of this relationship. In this paper, I try to determine why human capital matters.

[*]jmshapir@fas.harvard.edu

[†]I am grateful to Edward Glaeser, Claudia Goldin, Matthew Kahn, Kevin M. Murphy and Chris Rohlfs for helpful comments.

[1]See, for example, Glaeser, Scheinkman and Shleifer (1995), Simon (1998), Simon and Nardinelli (2002), Glaeser and Shapiro (2003), and Simon (2002).

As I show more formally in the next section, there are essentially three possible explanations for the relationship between human capital and city growth. The first is omitted variable bias: some feature or features of an area that are correlated with both human capital and employment growth have been left out of the regression. I devote relatively little attention to this theory, as past research has tended to find that including broad sets of controls does not eliminate the positive effects of human capital (Glaeser, Scheinkman and Shleifer, 1995; Glaeser and Shapiro, 2003).

The next hypothesis is that a highly educated population generates greater local productivity growth, perhaps through knowledge spillovers.[2] A number of researchers have adopted this explanation (see, for example, Simon and Nardinelli, 2002), and it has received some support from the work of Rauch (1993) and Moretti (2003), who show that, conditional on observable worker characteristics, wages are higher in high human capital cities.[3]

The final explanation is that areas with more educated populations experience more rapid growth in the quality of life. This might occur, say, because more educated individuals improve amenities in cities in which they reside, or because they seek out areas in which quality of life is rising.[4]

As I show in the next section, it is possible to use data on wage and land price growth to distinguish between the productivity and quality-of-life explanations. In a simple neoclassical model in which mobile firms bid for workers and mobile households bid for land, changes in wages and land prices will capitalize changes in local produc-

[2]Lucas (1988) discusses the role of knowledge spillovers in country and city growth. Black and Henderson (1999) develop a model of endogenous urban growth that embeds local effects of human capital accumulation

[3]By contrast, Acemoglu and Angrist (1999) find using an instrumental variables approach that the external effects of human capital at the state level are relatively small.

[4]As evidence on the latter possibility, Kahn (2000) documents that the reduction of ozone smog in San Bernardino was accompanied by in-migration of the college-educated. And Cullen and Levitt (1999) show that the migration decisions of better educated households are more sensitive to the level of crime in a city.

tivity and local consumption amenities. Using Census data from 1940 through 1990, I show that metropolitan areas richer in skilled residents tend to experience faster growth in both wages and house values, with the latter effect generally much larger than the former. These relationships hold after controlling for observable worker and house characteristics, so it seems plausible that they are not driven merely by changes in the composition of the labor and housing markets.

A calibration of a simple but fairly general city growth model suggests that roughly 63 percent of the effect of human capital on employment growth is due to productivity; the rest comes from the relationship between concentrations of skill and growth in the quality of life.

The organization of the paper is as follows. Section 1 presents a simple model of city growth and illustrates the three possible explanations for the relationship between human capital and metropolitan area employment growth. Section 2 describes the Census data I use to conduct the estimation. Section 3 presents evidence on the relationship between human capital and growth in employment, wages, and housing costs. Section 4 concludes.

1 Estimating framework

In this section I develop a simple neoclassical model of city growth, and use it to illustrate three hypotheses about the correlation between growth and human capital. The model is based on Roback's (1982) formulation, which has been used extensively to generate city-level rankings of quality of life and to infer the value to consumers and firms of various local public goods or city characteristics.[5] Most studies have exploited the cross-sectional implications of the Roback model; here I will place the

[5]See, for example, Blomquist, Berger and Hoehn (1988) Gyourko and Tracy (1991), Cragg and Kahn (1997), and Black (1999).

model in a more dynamic context.[6]

Before presenting the formal model, it will be helpful to discuss the intuition behind it. Consider a world of identical firms and households choosing among a set of locations. Each location is endowed with a productive amenity (that enters the production function) and a consumption amenity (that enters the utility function). Suppose that households consume only land and a traded good and that firms use only labor as an input. Let us first consider equilibrium in production, which requires that all firms be indifferent between locations. In equilibrium, wages must be higher in more productive locations, because otherwise firms would move into those locations and bid up the price of labor. In order for households to be indifferent between more and less productive locations, land prices must be higher in more productive places because wages will be higher in those locations. Land prices must also capitalize consumption amenities; that is, land will be more expensive in more pleasant locations.

These equilibrium conditions hold equally well in a dynamic context. If a city experiences relative growth in its productivity, then it should experience growth in both wages and land prices; if it experiences growth in quality of life, this will tend to be reflected in land price growth. In a more general model in which firms use land as an input to production, these equilibrium conditions must be modified somewhat, but it remains possible to identify changes in productive and consumption amenities using data on wages and land prices in a set of locations.

To see these results formally, consider an economy with a set of locations $i \in \{1, 2, ..., I\}$, each endowed with location specific productivity and quality of life, denoted A_i and Q_i, respectively. Firms produce a homogeneous good sold on a world market at the numeraire price of 1 using a constant returns to scale production func-

[6] The Roback model's implications for growth have been addressed before, however. For example, Glaeser, Scheinkman and Shleifer (1995) use a parametric example of the more general model to make inferences about the causes of city growth.

tion $Y = AF\left(L, R^f\right)$, where L denotes the quantity of labor and R^f the quantity of land used in production. Input markets are competitive, and firms face a constant per-unit marginal cost given by the function $\frac{C(W_i,P_i)}{A_i}$, where W_i and P_i are the prices of labor and land in location i. Spatial equilibrium requires that this marginal cost be equal to unity at all locations, so that our first equilibrium condition is given by

$$C\left(W_i, P_i\right) = A_i \tag{1}$$

for all i.

Consumers have preferences given by $U = U\left(Q, X, R^c\right)$, where X is the quantity of goods consumed and R^c is the quantity of land consumed. This utility function implies an indirect utility function $V\left(Q_i, W_i, P_i\right)$ which, in equilibrium, must be constant across locations. Our second condition is therefore

$$V\left(Q_i, W_i, P_i\right) = \overline{U} \tag{2}$$

for all i. To close the model, I will suppose that $P_i = f\left(L_i\right)$, with $f'\left(\cdot\right) > 0$, i.e. that there is an increasing supply price of housing.

Allow A_i and Q_i to change exogenously over time. We can totally differentiate equilibrium conditions (1) and (2) as follows:

$$\begin{aligned} C_W \frac{dW_i}{dt} + C_P \frac{dP_i}{dt} &= \frac{dA_i}{dt} \\ V_Q \frac{dQ_i}{dt} + V_W \frac{dW_i}{dt} + V_P \frac{dP_i}{dt} &= \frac{d\overline{U}}{dt}. \end{aligned} \tag{3}$$

Let k_R and k_L be the shares of land and labor in the firm's cost function, s_R be the share of land in the household's budget, and denote natural logarithms of variables with lowercase letters. I will normalize $\frac{d\overline{U}}{dt} = 0$. Then we can rearrange the above conditions to yield expressions for the changes in wages and land rents:

$$\begin{aligned} \frac{dp_i}{dt} &= \frac{1}{\frac{k_R}{k_L} + s_R}\left(\frac{V_Q Q}{V_W W}\frac{dq_i}{dt} + \frac{1}{k_L}\frac{da_i}{dt}\right) \\ \frac{dw_i}{dt} &= \frac{1}{k_L}\frac{s_R}{\frac{k_R}{k_L} + s_R}\frac{da_i}{dt} - \frac{\frac{k_R}{k_L}}{\frac{k_R}{k_L} + s_R}\frac{V_Q Q}{V_W W}\frac{dq_i}{dt}. \end{aligned} \tag{4}$$

Additionally, given the assumed supply curve of land, if we let σ be the elasticity of land rents with respect to local employment, employment growth can be written as

$$\frac{dl_i}{dt} = \frac{1}{\sigma} \frac{1}{\frac{k_R}{k_L} + s_R} \left(\frac{V_Q Q}{V_W W} \frac{dq_i}{dt} + \frac{1}{k_L} \frac{da_i}{dt} \right). \tag{5}$$

These conditions must hold for all cities i.

Changes in land rents will capitalize growth in productivity and in the quality of life, scaled by the importance of land in the firm and household budgets. Changes in wages will reflect productivity growth, less a correction to compensate firms for changes in land prices. In the limiting case in which firms use no land in the production process, wage growth will directly capitalize productivity growth. The above equations therefore suggest a framework for evaluating the extent to which quality of life and productivity growth are associated with a given correlate of employment growth.

To see this formally, let $H_{i,t}$ denote the concentration of human capital in city i at time t, and let $X_{i,t}$ be a vector of other city characteristics. Suppose that

$$\begin{aligned} \frac{V_Q Q}{V_W W} \Delta q_{i,t+1} &= H_{i,t} \beta^q + X_{i,t} \gamma^q + \epsilon^q_{i,t+1} \\ \Delta a_{i,t+1} &= H_{i,t} \beta^a + X_{i,t} \gamma^a + \epsilon^a_{i,t+1} \end{aligned} \tag{6}$$

where Δ denotes changes. Suppose further that the shocks ϵ^q and ϵ^a are drawn independently of X and H.[7] Then, by equation (5) above, we have

$$\Delta l_{i,t+1} = \frac{1}{\sigma} \frac{1}{\frac{k_R}{k_L} + s_R} \left(H_{i,t} \left(\beta^q + \frac{1}{k_L} \beta^a \right) + X_{i,t} \left(\gamma^q + \frac{1}{k_L} \gamma^a \right) \right) + \varepsilon^l_{i,t+1} \tag{7}$$

where

$$\varepsilon^l_{i,t+1} = \frac{1}{\sigma} \frac{1}{\frac{k_R}{k_L} + s_R} \left(\epsilon^q_{i,t+1} + \frac{1}{k_L} \epsilon^a_{i,t+1} \right). \tag{8}$$

[7] The shocks are not assumed to be identically distributed, however, nor are they assumed to be drawn independently over time or independently of one another. That is, I will allow for the possibility that ε^q and ε^a are heteroskedastic, serially correlated, and correlated with one another.

Suppose that a positive correlation is observed between human capital $H_{i,t}$ and subsequent employment growth $\Delta l_{i,t+1}$. Equation (7) illustrates the three possible, non-mutually exclusive explanations for such a correlation:

1. *Omitted variables bias.* A positive relationship between $H_{i,t}$ and $\Delta l_{i,t+1}$ could arise if $H_{i,t}$ is correlated with some omitted component of $X_{i,t}$, and that omitted city characteristic is itself a cause of rapid employment growth. For example, if high human capital individuals tend to concentrate in cities with more rapidly growing industries, and city growth is affected by the growth of local industries, a correlation between human capital and employment growth could arise.

2. *Productivity growth.* If high human capital is associated with more rapid productivity growth, that is, if $\beta^a > 0$, then human capital $H_{i,t}$ will be positively correlated with subsequent employment growth $\Delta l_{i,t+1}$.

3. *Growth in the quality of life.* Suppose that cities with higher concentrations of human capital experience faster growth in the quality of life, that is, suppose that $\beta^q > 0$. Then human capital and employment growth will covary positively.

The focus of this paper is on evaluating the relative importance of hypotheses (2) and (3). This requires estimating β^a and β^q, the parameters relating human capital to growth in productivity and quality of life, respectively. Suppose we have data (possibly noisy) on changes in land prices and wages for a panel of cities. Note that by (4) we can write

$$\begin{aligned}
\Delta p_{i,t+1} &= \frac{1}{\frac{k_R}{k_L} + s_R} \left(\frac{V_Q Q}{V_W W} \Delta q_{i,t+1} + \frac{1}{k_L} \Delta a_{i,t+1} \right) + \mu^p_{i,t+1} \qquad (9) \\
\Delta w_{i,t+1} &= \frac{1}{k_L} \frac{s_R}{\frac{k_R}{k_L} + s_R} \Delta a_{i,t+1} - \frac{\frac{k_R}{k_L}}{\frac{k_R}{k_L} + s_R} \frac{V_Q Q}{V_W W} \Delta q_{i,t+1} + \mu^w_{i,t+1}
\end{aligned}$$

where μ^p and μ^w are measurement error in price and wage growth, respectively, and are assumed to be independent of Δa and Δq.[8] Rearranging (9) we have that

$$\begin{aligned} k_L \Delta w_{i,t+1} + k_R \Delta p_{i,t+1} &= H_{i,t}\beta^a + X_{i,t}\gamma^a + \epsilon^a_{i,t+1} + k_L \mu^p_{i,t+1} + k_R \mu^w_{i,t+1} \quad (10) \\ s_R \Delta p_{i,t+1} - \Delta w_{i,t+1} &= H_{i,t}\beta^q + X_{i,t}\gamma^q + \epsilon^q_{i,t+1} + s_R \mu^p_{i,t+1} - \mu^w_{i,t+1}. \end{aligned}$$

Given values of k_L, k_R, s_R, it is thus possible to use data on growth in wages and land prices to determine the relative importance of productivity and quality of life in explaining the relationship between human capital and city employment growth.

2 Data description

To form the basic panel of metropolitan areas, I extracted from the IPUMS database (Ruggles and Sobeck, 1997) all prime-age (25 to 55) white males living in Census-defined metropolitan areas in the years 1940, 1970, 1980, and 1990. My measure of total employment in a given metropolitan area in a given year is a count of the total number of prime-age white males in the sample.[9] I construct an area-level employment growth measure for each time period as the log change in employment. I standardize this to be a ten-year growth rate in the 1940-1970 period.

I construct the wage series as follows. I restrict attention to white prime-age males living in metropolitan areas. To construct a wage estimate, I divide total wage and salary income for each individual by total annual hours worked, imputed from the categorical variables on weeks and hours worked available in the microdata.[10] I then regress the log of the wage for each individual on dummies for each metropolitan area, age and its square, and dummies for veteran status, marital status, educational

[8] As with ε^q and ε^a, it will not be necessary to assume that μ^p and μ^w are homoskedastic, independent over time, or drawn independently of one another.

[9] I have used person-level sample weights wherever appropriate in constructing my measures of employment, human capital, and other metropolitan area characteristics.

[10] In all cases I used the midpoint of the categorical range as the point estimate.

attainment, industry category and occupational category.[11] All regressions include dummies for missing values of marital and veteran status; observations with missing values of other variables were dropped. These regressions were run separately for each Census year so as to avoid unnecessary restrictions on the coefficients.

For each year I extract the coefficients on the metropolitan area dummies to be used as estimates of local differences in wages.[12] Naturally, these estimates are only as good as the controls–sorting on omitted characteristics will introduce bias. However, as table 1 illustrates, the estimates generally seem sensible. Moreover, for the purposes of studying growth the changes in these residuals are more important than their levels–and growth rates in wage residuals will at least be purged of time-invariant differences in the characteristics of local workers.

To construct the house value series I employ a similar procedure using Census data on reported house values for owner-occupied units. Unfortunately these data are not available for 1950, and data on housing characteristics are not available for 1940. Therefore when using data on house values I will generally report results both with and without the 1940-1970 time period . I run house value regressions separately within each year. The controls for housing characteristics I employ are dummies for commercial use status, acreage of property, availability of kitchen or cooking facilities, number of rooms, type of plumbing, year built, number of units in structure, water source, type of sewage disposal, and number of bedrooms.[13] These controls, however incomplete, were available for all years (except 1940) and therefore permit me to construct a consistent series. Again, while bias due to omitted housing characteristics may be a problem, it may be less of a concern when using growth rates than when using levels. Moreover, table 1 suggests that my estimates of metropolitan

[11] Further details on the controls used are available in subsection 1 of the Appendix.

[12] The use of metropolitan area dummies to measure local wage and price differences is related to the approach taken in Gabriel and Rosenthal (2003).

[13] Subsection 2 of the Appendix contains additional details about the controls used.

area level house value effects are sensible.

As a measure of the concentration of human capital in a metropolitan area, I calculate the sample share of prime-age white males who fall into each of the following categories: high school degree only, some college, and college degree or higher. Appendix table 1 presents summary statistics for these shares by time period.

3 Results

Table 2 reports coefficients from ordinary least squares (OLS) regressions of employment, wage and house value, growth on the log of the percent college graduates for various time periods. Data on wages and house values come from metropolitan area fixed effects in hedonic regressions of prices on worker or housing unit characteristics as described in the previous section, and are therefore purged of observable differences in worker or housing unit quality. In regressions that include multiple time periods, dummies for time period are included, and standard errors are adjusted for correlation of the errors within metropolitan areas.

These regressions reveal a number of important facts. First, they confirm the usual finding that cities with greater concentrations of human capital experience more rapid growth in employment. A 10 percent increase in the share of college educated residents is associated with an increase in the employment growth rate of roughly .6 percent in most specifications. In the 1970-90 and 1980-90 samples the coefficient is not statistically significant, although the standard errors are too large in these cases to rule out a substantial effect.

A second important pattern is that growth in wages and house values tend to be higher in cities with greater concentrations of college-educated residents. The only exception is the 1970-80 period, in which the wage growth coefficient is negative. In the overall (1940-90) sample, a 10 percent increase in the share of college educated residents corresponds to a .2 percent increase in wage growth and a .7 percent increase in the growth of house values, both statistically significant. In general, these effects seem to increase over time, perhaps reflecting the rising importance of college (rather than high school) education.

A final observation on table 2 is that in all samples the effect of human capital on house value growth exceeds the effect on wage growth. In the overall sample (1940-90), the effect of the log of the share college educated on growth in house values is more than three times as large as the effect on growth in wages.

The reduced-form facts presented in table 2 suggest that growth in quality of life may be playing an important role in the relationship between human capital and growth, since growth in house values seems generally to be more sensitive to the share of college educated residents than growth in wages. For a more quantitative evaluation of the relative importance of quality of life and productivity in explaining the human capital-growth relationship, we will need to estimate equations (10). For this we require values for labor's share of output (k_L), land's share of output (k_R), and the share of land in the household budget (s_R).

Krueger (1999) estimates that labor's total share of output (including the return to human capital) is roughly .75; Poterba (1997) also places it at between 70 and 80 percent of national income. I will therefore use $k_L = .75$. Poterba (1997) reports a corporate capital income share of around 10 percent, placing an upper bound of around .15 on k_R. I will set $k_R = .10$, which is close to the upper bound and if anything seems likely to cause me to overstate the productivity effects of human capital.

The literature has traditionally used a value of about .05 for s_R, which derives from an effort to account for the typical household's expenditure on land (Roback, 1982). In principle, that is the quantity demanded by theory, but in practice this estimate is likely to be far too small. The reason is that the model in section 1 assumes that all goods other than land are traded on a national market and therefore display no local price variation. In a more realistic framework, s_R is not merely the household budget share of land per se but rather the share in the household budget of all goods that are produced using local land as an input. In other words, s_R should capture the importance of all "cost of living" differences between locations, because all of these costs matter in equilibrating population across cities. Using this logic, I show in subsection 3 of the Appendix that reasonable values of s_R are likely to be in the vicinity of .5. However, I will report results for a wide range of values to permit flexibility in interpreting my findings.

As I showed in section 1, regressions of $k_L \Delta w_{i,t+1} + k_R \Delta p_{i,t+1}$ and $s_R \Delta p_{i,t+1} - \Delta w_{i,t+1}$ on the log of the share of college graduates will yield estimates of the parameters β^a and β^q. These estimates, denoted $\hat{\beta}^a$ and $\hat{\beta}^q$, capture the effect of human capital on growth in productivity and the quality of life, respectively. Since the total effect of human capital on employment growth is equal to $\beta^q + \frac{1}{k_L}\beta^a$ (see equation (7)), the fraction of the employment growth effect that is due to productivity growth can be estimated as $\frac{\frac{1}{k_L}\hat{\beta}^a}{\hat{\beta}^q + \frac{1}{k_L}\hat{\beta}^a}$.

Table 3 shows the results of this exercise for the whole (1940-90) sample. While results vary with the choice of s_R, at my preferred value of .5 roughly 63 percent of the overall growth effect of human capital is attributed to productivity growth. This suggests that while knowledge spillovers do play an important role in the growth effects of human capital, consumption amenities are an important component as well. Even for $s_R = 0.4$, over one fourth of the total growth effect is attributed to growth in local quality of life.

Overall, then, my findings indicate an important role of quality of life in driving the relationship between the share of college-educated residents in a metropolitan area and the area's subsequent employment growth. While the literature has tended to emphasize productive externalities from human capital, this evidence suggests there may be important consumption externalities as well.

4 Conclusions

Several possible mechanisms might underlie the relationship between the concentration of skilled residents in a metropolitan area and subsequent growth in the area's quality of life.

First, skilled residents may be the first to flee areas experiencing declines in consumption amenities and the first to enter areas experiencing improvements. This mechanism is consistent with Kahn (2000), who finds that college-educated residents are more likely to move into an area in response to a reduction in smog, and Cullen and Levitt (1999), who show that the migration decisions of high-skilled households are more sensitive to the level of crime in a city.

Second, concentrations of skilled residents may encourage the growth of consumer markets, such as restaurants and bars, which then make an area more attractive to potential migrants. In line with this hypothesis, Glaeser, Kolko and Saiz (2001) show evidence that cities with superior markets for goods and services experience more rapid population growth.

Third, highly educated households may act, through the political system or privately, to improve local quality of life, perhaps because of a desire to raise property values. Moreover, better educated households are more likely to be homeowners, and some evidence exists to suggest that homeowners make greater investments in their local communities (Glaeser and Shapiro, forthcoming).

While these hypotheses hardly constitute an exhaustive list of possible explanations, they do suggest a number of paths for future research to uncover why local areas with greater concentrations of skill seem to experience more rapid growth in consumption amenities.

References

[1] Daron Acemoglu and Joshua Angrist. How large are the social returns to education? evidence from compulsory schooling laws. *National Bureau of Economic Research Working Paper*, 7444, December 1999.

[2] Duncan Black and Vernon Henderson. A theory of urban growth. *Journal of Political Economy*, 107(2):252–284, April 1999.

[3] Sandra E. Black. Do better schools matter? parental valuation of elementary education. *Quarterly Journal of Economics*, 114(2):577–599, May 1999.

[4] Glenn C. Blomquist, Mark C. Berger, and John P. Hoehn. New estimates of quality of life in urban areas. *American Economic Review*, 78(1):89–107, March 1988.

[5] Michael Cragg and Matthew E. Kahn. New estimates of climate demand: Evidence from location choice. *Journal of Urban Economics*, 42:261–284, 1997 1997.

[6] Julie Berry Cullen and Steven D. Levitt. Crime, urban flight, and the consequences for cities. *Review of Economics and Statistics*, 81(2):159–169, May 1999.

[7] Stuart A. Gabriel and Stuart F. Rosenthal. Quality of the business environment versus quality of life: Do firms and households like the same cities? *Review of Economics and Statistics*, February 2003. forthcoming.

[8] Edward L. Glaeser, Jed Kolko, and Albert Saiz. Consumer city. *Journal of Economic Geography*, 1(1):27–50, January 2001.

[9] Edward L. Glaeser, Jose A. Scheinkman, and Andrei Shleifer. Economic growth in a cross-section of cities. *Journal of Monetary Economics*, 36:117–143, 1995.

[10] Edward L. Glaeser and Jesse M. Shapiro. The benefits of the home mortgage interest deduction. *Tax Policy and the Economy*, 17, September 2003. Forthcoming.

[11] Edward L. Glaeser and Jesse M. Shapiro. Urban growth in the 1990s: Is city living back? *Journal of Regional Science*, 43(1):139–165, February 2003.

[12] Joseph Gyourko and Joseph Tracy. The structure of local public finance and the quality of life. *Journal of Political Economy*, 99(4):774–806, August 1991.

[13] Matthew E. Kahn. Smog reduction's impact on california county growth. *Journal of Regional Science*, 40(3):565–582, August 2000.

[14] Alan B. Krueger. Measuring labor's share. *American Economic Review*, 89(2):45–51, May 1999.

[15] Robert E. Lucas. On the mechanics of economic development. *Journal of Monetary Economics*, 22:3–42, 1988.

[16] Enrico Moretti. Human capital externalities in cities. *National Bureau of Economic Research Working Paper*, 9641, April 2003.

[17] James Poterba. The rate of return to corporate capital and factor shares: New estimates using revised national income accounts and capital stock data. *National Bureau of Economic Research Working Paper*, 6264, 1997.

[18] James E. Rauch. Productivity gains from geographic concentration of human capital: evidence from the cities. *Journal of Urban Economics*, 34:380–400, 1993.

[19] Jennifer Roback. Wages, rents, and the quality of life. *Journal of Political Economy*, 90(6):1257–1278, December 1982.

[20] Stephen Ruggles and Matthew Sobeck. *Integrated public use microdata series: Version 2.0.* Historical Census Projects, University of Minnesota, Minneapolis, 1997.

[21] Curtis J. Simon. Human capital and metropolitan employment growth. *Journal of Urban Economics*, 43:223–243, March 1998.

[22] Curtis J. Simon. Industrial reallocation across u.s. cities, 1977-97: The role of human capital. *Clemson University Mimeograph*, August 2002.

[23] Curtis J. Simon and Clark Nardinelli. Human capital and the rise of american cities, 1900-1990. *Regional Science and Urban Economics*, 32:59–96, January 2002.

Table 1: Highest and lowest wage and house value fixed effects, 1990

A. Wage fixed effects

Highest	Stamford, CT	0.60
	Norwalk, CT	0.55
	Danbury, CT	0.41
	New York-Northeastern NJ	0.39
	Bridgeport, CT	0.38
Lowest	Alexandria, LA	-0.11
	Laredo, TX	-0.12
	Bryan-College Station, TX	-0.13
	McAllen-Edinburg-Pharr-Mission, TX	-0.18
	Brownsville - Harlingen-San Benito, TX	-0.22

B. House value fixed effects

Highest	San Jose, CA	1.21
	Honolulu, HI	1.19
	Stamford, CT	1.17
	Santa Cruz, CA	1.15
	Norwalk, CT	1.09
Lowest	McAllen-Edinburg-Pharr-Mission, TX	-0.60
	Sioux City, IA/NE	-0.60
	Flint, MI	-0.62
	Joplin, MO	-0.63
	Johnstown, PA	-0.64

Notes: Wage fixed effects reflect coefficients from metropolitan area dummies in a regression of log(wage) on these dummies and controls for observable worker characteristics. House value fixed effects reflect coefficients from metropolitan area dummies in a regression of log(house value) on these dummies and controls for observable housing characteristics. See section 2 of text for details.

Table 2: Human capital and growth

Independent variable: log(share college educated)

Sample	Number of observations	Dependent variable is growth in... Employment	Wage	House value
1940-90	498	0.0647	0.0197	0.0745
		(0.0245)	(0.0080)	(0.0189)
1970-90	369	0.0621	0.0382	0.1364
		(0.0450)	(0.0143)	(0.0372)
1940-70	129	0.0669	0.0040	0.0223
		(0.0229)	(0.0087)	(0.0126)
1970-80	117	0.1257	-0.0223	0.1207
		(0.0630)	(0.0170)	(0.0486)
1980-90	252	0.0334	0.0654	0.1434
		(0.0555)	(0.0201)	(0.0477)

Notes: Table shows coefficient in regression of dependent variable on the log of the percent of prime-age white males with a college degree in the metropolitan area. Wage and house value growth are measured as the growth in metropolitan area fixed effects from hedonic regressions as described in section 2 of the text. Regressions include time period dummies where appropriate. Standard errors have been adjusted for serial correlation within metropolitan areas where appropriate. All standard errors are heteroskedasticity-robust.

Table 3: Human capital and growth in productivity and the quality of life

Independent variable: log(share college educated)

s_R	Dependent variable		Productivity share of growth effect
	$k_L \Delta w_{i,t+1} + k_R \Delta p_{i,t+1}$	$s_R \Delta p_{i,t+1} - \Delta w_{i,t+1}$	$\frac{\frac{1}{k_L}\hat{\beta}^a}{\hat{\beta}^q + \frac{1}{k_L}\hat{\beta}^a}$
.05	0.0222 (0.0069)	-0.0159 (0.0077)	2.16
.25		-0.0010 (0.0075)	1.03
.4		0.0102 (0.0086)	0.74
.5		**0.0176** **(0.0097)**	**0.63**
.6		0.0251 (0.0110)	0.54
.75		0.0363 (0.0133)	0.45
1		0.0549 (0.0174)	0.35

Notes: Table shows coefficient in regression of dependent variable on the log of the share of prime age white males in the metropolitan area with a college degree. All calculations use $k_L = .75$, $k_R = .10$. I measure $\Delta w_{i,t+1}$ as the change in a metropolitan area i's log(wage) fixed effect from time t to $t+1$, as described in section 2; $\Delta p_{i,t+1}$ is measured similarly using data on house values. All regressions include time period dummies. All standard errors have been adjusted for serial correlation within metropolitan areas.

5 Appendix

5.1 Measuring Local Area Wages

In order to measure relative wage levels in metropolitan areas at time t, I regress the log wage of all prime-age males in the sample at time t on dummies for metropolitan areas and a set of controls. These controls are age in years, the square of age in years, and dummies for the following worker characteristics (IPUMS variable name in parentheses):

- Veteran status (VETSTAT): The veteran status categories are not applicable (code 1); no service (code 2); yes (code 3); and not ascertained (code 4).

- Marital status (MARST): The marital status categories used are based on current marital status. The categories are married, spouse present (code 1); married, spouse absent (code 2); separated (code 3); divorced (code 4); widowed (code 5); never married, single, or not applicable (code 6).

- Educational attainment (EDUCREC): The education categories used are based on the educational attainment recode. The categories, which correspond to completed years of schooling, are none or preschool (code 1); grade 1, 2, 3, or 4 (code 2); grade 5, 6, 7, or 8 (code 3); grade9 (code 4); grade 10 (code 5); grade 11 (code 6); grade 12 (code 7); 1, 2, or 3 years of college (code 8); 4+ years of college (code 9). Observations with missing data on educational attainment were dropped from the wage regression.

- Occupation category (OCC1950): Occupational categories are based on the 1950 classification. The categories are professional and technical (codes 000-099); farmers (100-199); managers, officials, and proprietors (200-299); clerical and kindred (300-399); sales workers (400-499); craftsmen (500-599); operatives (600-699); service (700-799); farm laborers (800-899); laborers (900-970). Observations with missing data on occupation were dropped from the wage regression.

- Industry category (IND1950): The industry categories I use are based on the 1950 industrial classification. They are agriculture, forestry, and fishing (codes 105-126); mining (206-236); construction (246); durable goods manufacturing (300-399); nondurable goods manufacturing (400-499); transportation (506-568); telecommunications (578-579); utilities and sanitary services (586-598); wholesale trade (606-627); retail trade (636-699); finance, insurance, and real estate (716-756); business and repair services (806-817); personal services (826-849); entertainment and recreation services (856-859); professional and related services (868-899); and public administration (900-936). Observations with missing data on industry were dropped from the regression.

5.2 Measuring Local Area House Values

My housing dataset consists of all households not residing in group quarters. In order to measure relative house values in metropolitan areas in 1970, 1980, and 1990, I regress the log reported value of all owner-occupied houses in the sample in each year on dummies for metropolitan areas and a set of controls. For 1940, the controls are not available so the regression includes only the metropolitan area dummies. The controls used in the 1970, 1980, and 1990 samples are dummies for the following housing characteristics (IPUMS variable name in parentheses):

- Commercial use status (COMMUSE): The commercial use status categories allow identification of owner-occupied homes attached to businesses or medical/dental offices. The categories are not applicable (code 0); no commercial use (code 1); commercial use (code 2); and unknown, unit on 10+ acres (code 3, 1970 only).

- Acreage of property (ACREPROP): This variable indicates whether a non-city, non-suburban unit is on 10 or more acres. The categories are city or suburban lot (code 1, 1970 only); city or suburban lot or rural lot less than 1 acre (code 2, 1980 and 1990); non-city, non-suburban lot under 10 acres including less than 1 acre (code 3, 1970 only); non-city, non-suburban lot 1-9 acres (code 4, 1980 and 1990); non-city, non-suburban lot 10+ acres (code 5, 1980 and 1990).

- Availability of kitchen or cooking facilities (KITCHEN): This variable indicates whether a housing unit has a kitchen, defined as a sink with piped water, a nonportable cooking device, and an electronic refrigerator. The categories are not applicable (code 0); no kitchen (code 1); shared use kitchen (code 3, 1970 only); shared or exclusive use kitchen (code 4, 1980 and 1990); exclusive use kitchen (code 5, 1970 only).

- Number of rooms (ROOMS): This variable indicates the number of whole rooms in the housing unit. The categories are not applicable (code 0), one room (code 1), two rooms (code 2), etc., with a top-code at 9 rooms (code 9).

- Type of plumbing (PLUMBING): This variable indicates whether the housing unit has complete plumbing facilities and, in some years, the nature of any partial facilities. The categories are not applicable (code 0), lacking complete plumbing (code 10, 1990 only), lacking hot water (code 11, 1970 only), lacking other or all plumbing facilities (code 12, 1970 only), has some facilities (code 13, 1980 only), has no facilities (code 14, 1980 only), complete plumbing (code 20, 1970 and 1990), exclusive use complete plumbing (code 21, 1980 only), shared complete plumbing (code 22, 1980 only).

- Year built (BUILTYR): This variable codes the age of the structure in years. The categories are not applicable (code 0), 0-1 year old (code 1), 2-5 years (code 2), 6-10 years (code 3), 11-20 years (code 4), 21-30 years (code 5), 31-40 years (code 6, 31+ in 1970), 41-50 years (code 7, 1980 and 1990, 41+ in 1980), 51+ years (code 8, 1990 only).

- Number of units in structure (UNITSSTR): Codes the number of occupied or vacant units in the structure. Categories are not applicable (code 0); mobile home or trailer (code 1); boat, tent, van, other (code 2); single-family detached (code 3); single-family attached (code 4); two-family building (code 5); 3-4 family building (code 6); 5-9 family building (code 7); 10-19 family building (code 8); 20-49 family building (code 9); 50+ family building (code 10).

- Water source (WATERSRC): Categories are not applicable (code 0); public system or private company (code 1); individual well (code 2); individual well, drilled (code 3); individual well, dug (code 4); other source (code 5).

- Type of sewage disposal (SEWAGE): Categories are not applicable (code 0); public sewer (code 1); septic tank or cesspool (code 2); other means (code 3).

- Number of bedrooms (BEDROOMS): Categories are not applicable (code 0), no bedrooms (code 1), 1 bedroom (code 2), 2 bedrooms (code 3), 3 bedrooms (code 4),4 bedrooms (code 5), 5+ bedrooms (code 6).

5.3 Calibrating the Household Share of Land (s_R)

The share of land in the budget, s_R, ought to capture the share of household expenditures that go to nontradeable goods. That is, it should reflect how "cost of living" varies with land prices across cities. The literature has typically fixed this parameter at roughly .05, a value intended to reflect the literal share of land in the household's budget. While there is no perfect way to calculate the true value, I will argue in this section that it is likely to be on the order of .5.

ACCRA (www.accra.org) compiles data on cost-of-living differences between U.S. cities, both overall and for specific categories of goods. Appendix table 2 shows the composite and grocery price indices for the 19 of the top 20 cities from the third quarter of 1999. Cities are put in descending order by my measure of house values for the corresponding metropolitan area. As the table makes clear, the composite price index–meant to capture all cost-of-living differences between locations–varies strongly with the house value measure. Moreover, the prices of groceries, which in principle are a highly tradeable good, vary considerably with underlying land prices.

A regression of the log of the composite index on the house value measure yields a coefficient of .35; removing New York City (an outlier) brings this down to .26. That is, a one percent increase in the price of land corresponds to an increase in the overall cost of living of between .26 and .35 percent. It seems therefore that the value of s_R is likely to be considerably larger than .05.

As a further justification for using values of s_R in the vicinity of .5, we can take advantage of the fact that weather (as measured by mean January temperature) has been a robust positive predictor of growth over the latter half of the twentieth century (Glaeser, Scheinkman and Shleifer, 1995; Glaeser and Shapiro, forthcoming). Since weather is presumably influencing growth more through quality of life than through productivity, studying how much of the weather effect is attributed to quality of life for different values of s_R will allow us to check the plausibility of various assumptions.

I repeated the exercise of section 3 using log of mean January temperature rather than log of the share of college graduates as the key independent variable. That is, I use data on wages and house values from the overall sample (1940-90 period) to calculate the effect of log of mean January temperature on growth in productivity and the quality of life over this period (regressions not shown). For $s_R = .05$, the model indicates that January temperature is *negatively* related to growth in the quality of life over this period. For $s_R = .25$, the model attributes roughly 40 percent of the effect of temperature on growth to quality of life, still attributing a majority of the temperature effect to productivity. For $s_R = .5$ and $s_R = .75$, I calculate that quality of life accounts for 64 and 74 percent of the overall growth effect, respectively.

This exercise indicates that the parameterization common in the literature over-attributes the growth effect of mean January temperature to productivity growth, whereas values of s_R in the vicinity of .5 attribute most of the effect to quality of life, consistent with *a priori* intuition about the causes of the weather effect.

Overall, then, the evidence seems consistent with a value of s_R on the order of .5, and quite inconsistent with values in the vicinity of .05. In section 3, I will report results for a range of values.

Appendix Table 1: Summary statistics for human capital measures

Time period	Number of cities	Mean share of		
		High school graduates	Some college	College graduates
1940	129	18.5	8.2	8.1
1970	117	35.1	14.2	18.3
1980	252	35.2	20.5	26.7

Means reflect sample shares of prime-age (25-55) white males in each category, averaged over all metropolitan areas.

Appendix Table 2: Cost of living differences between cities

Metropolitan area	House value fixed effect, 1990	ACCRA price index Composite	ACCRA price index Grocery
San Francisco-Oakland-Vallejo, CA	1.0806	156.6	121.6
Los Angeles-Long Beach, CA	1.0799	123.1	110.0
New York-Northeastern NJ	0.8674	231.8	141.5
Boston, MA	0.8452	136.2	114.5
San Diego, CA	0.8075	126.4	122.5
Washington, DC/MD/VA	0.5951	137.8	110.2
Seattle-Everett, WA	0.4735	118.7	109.7
Riverside-San Bernadino, CA	0.4637	114.7	114.4
Chicago-Gary-Lake IL	0.2563	109.0	109.3
Philadelphia, PA/NJ	0.1739	116.9	107.0
Baltimore, MD	0.1736	97.0	97.2
Minneapolis-St. Paul, MN	0.0546	105.4	99.5
Phoenix, AZ	0.0368	102.3	104.9
Atlanta, GA	-0.0308	103.2	106.7
Dallas-Fort Worth, TX	-0.0678	101.1	99.2
Cleveland, OH	-0.0974	112.2	108.9
St. Louis, MO/IL	-0.1228	97.3	99.6
Detroit, MI	-0.2199	112.9	106.2
Houston-Brazoria, TX	-0.2762	94.5	93.3

House value fixed effect column reports metropolitan area fixed effects from a regression of log of house value on housing unit characteristics, as described in section 2. ACCRA price indices are from www.accra.org and correspond to the third quarter of 1999. The price indices are normalized to have an average of 100 across all cities.