



# Human Capital Growth in a Cross Section of U.S. Metropolitan Areas

Christopher H. Wheeler

Growth of human capital, defined as the change in the fraction of a metropolitan area's labor force with a bachelor's degree, is typically viewed as generating a number of desirable outcomes, including economic growth. Yet, in spite of its importance, few empirical studies have explored why some economies accumulate more human capital than others. This paper attempts to do so using a sample of more than 200 metropolitan areas in the United States over the years 1980, 1990, and 2000. The results reveal two consistently significant correlates of human capital growth: population and the existing stock of college-educated labor. Given that population growth and human capital growth are both positively associated with education, these results suggest that the geographic distributions of population and human capital should have become more concentrated in recent decades. That is, larger, more-educated metropolitan areas should have exhibited the fastest rates of increase in both population and education and thus "pulled away" from smaller, less-educated metropolitan areas. The evidence largely supports this conclusion.

Federal Reserve Bank of St. Louis *Review*, March/April 2006, 88(2), pp. 113-32.

**H**uman capital is now commonly held to be one of the fundamental drivers of economic growth. To be sure, the notion that the skills possessed by an economy's workforce promote technological advancement and productivity growth is an intuitively appealing one. Yet, there is also a fair amount of empirical evidence that supports this notion. In particular, a sizable literature in the past two decades has established a strong statistical association between human capital (usually captured by educational attainment) and the growth of employment, productivity, and income. Moreover, this relationship holds with striking regularity at different levels of geographic aggregation, including countries (Barro, 1991), U.S. states (Barro and Sala-i-Martin, 1992), and cities and metropolitan areas (Glaeser, Scheinkman, and Shleifer, 1995; Glaeser and Saiz, 2003; and Simon and Nardinelli, 2002).

Economic growth, however, is only one benefit that has been associated with human capital. A variety of studies also suggest that greater educational attainment within local economies (e.g., states or cities) may tend to be accompanied by lower rates of crime (Lochner and Moretti, 2004), greater civic involvement (Dee, 2004; Milligan, Moretti, and Oreopoulos, 2004), and less political corruption (Glaeser and Saks, 2004). Clearly, because these are desirable outcomes, identifying the determinants of human capital growth is a worthwhile undertaking. Unfortunately, while a host of theoretical models have done so,<sup>1</sup> surprisingly little empirical research has followed suit. Most existing studies have focused on what human capital produces rather than why some economies accumulate more of

---

<sup>1</sup> See Barro and Sala-i-Martin (1995) for a survey of human capital-based models of growth.

---

Christopher H. Wheeler is a senior economist at the Federal Reserve Bank of St. Louis. Elizabeth La Jeunesse provided research assistance.

© 2006, The Federal Reserve Bank of St. Louis. Articles may be reprinted, reproduced, published, distributed, displayed, and transmitted in their entirety if copyright notice, author name(s), and full citation are included. Abstracts, synopses, and other derivative works may be made only with prior written permission of the Federal Reserve Bank of St. Louis.

## Wheeler

it than others.<sup>2</sup> As such, our understanding of human capital accumulation remains limited.

This paper looks at the growth of human capital in a sample of more than 200 U.S. metropolitan areas identified in the decennial U.S. Census over the years 1980, 1990, and 2000. Defining human capital accumulation as the change in the fraction of a metropolitan area's employed labor force with a bachelor's degree or more, I find that metropolitan areas with larger populations and higher fractions of their workers with a bachelor's degree tend to accumulate human capital at faster rates than less-populous, less-educated metropolitan areas. The results suggest that a 1-standard-deviation increase in either total resident population or the fraction of workers with a four-year college degree (in the cross section of metropolitan areas) tends to be associated with a 0.4- to 0.7-percentage-point rise in the share of college graduates in the workforce over the next decade. These estimated magnitudes, it should be noted, are *not* meant to be interpreted as causal, but simply to quantify the strength of the observed associations between these two variables and the accumulation of highly educated workers. Although some evidence suggests that certain measures of industrial composition and observable city-level amenities (e.g., restaurants and universities) are also associated with changes in the college fraction, none are as robustly correlated as population and the existing level of human capital.

These findings are intriguing, as they seem to suggest that the geographic distribution of human capital across the cities of the United States should have grown more concentrated (or unequal) between 1980 and 2000. After all, because human capital accumulation tends to be positively associated with the current level of human capital, the gap between initially high-education cities and low-education cities ought

to have widened in recent decades. The evidence strongly supports this conclusion. Various measures that characterize the degree of spread in the distribution of metropolitan area-level college attainment show rising dispersion between 1980 and 2000.

In addition, because previous research has established a positive link between population growth and education (e.g., Glaeser, Scheinkman, and Shleifer, 1995), one would expect to find a similar pattern of "divergence" in population levels across U.S. metropolitan areas in recent decades. That is, if more-populous cities accumulate highly educated workers more quickly than less-populous ones, then they should also gain population faster too. Rising educational attainment fuels population growth, which, in turn, spurs human capital accumulation and so on. This conclusion is also largely borne out in the data. The distribution of the logarithm of population became more concentrated within particularly large metropolitan areas between 1980 and 2000.

Although one might surmise that rising concentrations of population and education in the largest and most-educated cities have also led to a greater concentration of income, the evidence on this issue is somewhat mixed. In particular, while the data show that the distribution of metropolitan area-level average log hourly wages grew wider between 1980 and 1990, they also show that it narrowed slightly between 1990 and 2000. Growing concentrations of population and college-educated workers in the metropolitan areas of greatest size and abundance of human capital, then, have not been accompanied by substantial increases in the degree of inter-city (average) earnings inequality.

## DATA

The data used in the analysis are taken primarily from the 5 percent public use samples of the 1980, 1990, and 2000 U.S. Census as reported by the Integrated Public Use Microdata Series (Ruggles et al., 2004). These data files include a variety of personal characteristics, including age, education, and earnings, for samples of more

<sup>2</sup> There are two notable exceptions: Moretti (2004) offers a short analysis of the determinants of changing college attainment rates among U.S. metro areas, similar to what I do here. Glaeser and Saiz (2003) examine whether educational attainment responds to economic growth. With both of these papers, however, the primary issue under consideration is *not* the determinants of human capital growth. Consequently, their analyses are much more cursory with respect to this issue than my analysis here.

than 11 million individuals in each year, as well as information about each individual's place of residence. These data are used to construct a time series of metro area-level characteristics, including human capital.

In principle, "human capital" could be defined in many different ways: e.g., time spent on a particular job, time spent working on all jobs, numbers of different jobs held, educational attainment, some measure of "innate" ability or productivity. This paper takes a standard approach by using educational attainment, which can be justified by noting that (i) schooling has been shown to have a significant causal influence on individual productivity, at least as quantified by earnings (Card, 1999), and (ii) it tends to be strongly correlated with a variety of outcomes commonly theorized to be tied to "human capital," including economic growth. For these reasons, education is treated as a suitable metric for human capital. More specifically, I use the fraction of a metro area's employed labor force with a bachelor's degree or more because previous work on economic growth and education externalities in cities has found this particular quantity to capture variation in educational attainment reasonably well.<sup>3</sup>

Formally, metro areas in the analysis represent either metropolitan statistical areas (MSAs), New England county metropolitan areas (NECMAs), or consolidated metropolitan statistical areas (CMSAs) in the event that an MSA or NECMA belongs to a CMSA.<sup>4</sup> A total of 210 of these local markets are identified in the 1980 data, 206 in 1990, and 245 in 2000. Only 188 appear in all three Census years.

Additional characteristics describing metro areas are derived from the USA Counties CD-ROM (U.S. Bureau of the Census, 1999) and from County Business Patterns (CBP) files for the years 1980, 1990, and 2000. The former dataset provides information about county-level population and land area, which is used to generate population and population density figures at the metro-area

level.<sup>5</sup> The latter reports the numbers of various types of private sector establishments (e.g., restaurants and bars), which are used to characterize the amenity value of a metro area. Further details about the data appear in the appendix.

## EMPIRICAL FINDINGS

### *Human Capital and Urban Agglomeration*

Within the United States, human capital has typically been concentrated in metro areas. Among workers in the Census samples used here, 86.1 percent of all college graduates resided in a metro area in 1980. By 2000, this figure had risen to 89.9 percent. In contrast, approximately 78 percent of workers with only a high school diploma were metro dwellers in either year.

Why are highly educated workers drawn to cities? Numerous characteristics, of course, distinguish metro areas from non-metro areas and, thus, could offer some semblance of an answer. Besides larger and better-educated populations, urban agglomerations also tend to possess greater numbers of industries that highly educated workers may find particularly appropriate or appealing given their skills (e.g., professional and technical services). Metro areas also tend to offer a greater array of amenities (e.g., restaurants and museums), which may serve to attract and maintain a pool of highly educated labor (see Glaeser, Kolko, and Saiz, 2001).

Economically, the estimated returns to education do tend to be particularly high in metro areas. Consider, for instance, the results from a regression of log hourly earnings on five educational attainment indicators (no high school, some high school, a high school diploma only, some college or an associate's degree, a bachelor's degree or more), eight indicators representing years of potential work experience,<sup>6</sup> a metro residence dummy,

<sup>3</sup> See, for example, Black and Henderson (1999) and Moretti (2004).

<sup>4</sup> Throughout the paper, I use the terms "metropolitan area" and "city" interchangeably for expositional purposes. In all cases, "local markets" refers to MSAs, NECMAs, or CMSAs.

<sup>5</sup> County-level population data for the year 2000 are derived from the population estimates program of the U.S. Census Bureau at [www.census.gov/popest/estimates.php](http://www.census.gov/popest/estimates.php). In all years, land area from 1990 is used to compute density.

<sup>6</sup> These indicators represent 6-10 years of experience, 11-15 years, 16-20 years, 21-25 years, 26-30 years, 31-35 years, 36-40 years, and 41 or more years.

## Wheeler

and interactions between metro residence and each of the education and experience variables.<sup>7</sup> To keep the analysis simple, I have limited the sample of workers used for this regression to white males between the ages of 18 and 65. I have also performed the estimation separately for the 1980 and 2000 samples to account for any changes in the coefficient values over time.<sup>8</sup>

The resulting coefficient estimates, which for the sake of conciseness have been limited to the education variables, appear in Table 1. The raw coefficients on the five educational attainment dummies in the first five rows of results can be interpreted as the average log wages (conditional on all of the other covariates in the model) for workers in these education groups who reside *outside* of a metro area. The average log wages for workers inside metro areas is then given by the sum of these raw coefficients and the corresponding interaction listed in the remaining rows of the table.

With this interpretation in mind, it is evident that, although college graduates earn more than workers with less schooling, the premium associated with a college degree is particularly high within metro areas. In non-urban areas in 1980, for example, college-educated workers earned approximately 30 percent more than workers with only a high school diploma.<sup>9</sup> Within metro areas, that differential was 45 percent. By the year 2000, the college premium had risen to 49 percent outside of metro areas, 75 percent within them. In terms of raw (conditional) wage levels, college graduates earned an average of \$10.48 per hour outside of metro areas in 1980, \$12.26 within them.<sup>10</sup> By 2000, these figures stood at, respectively, \$10.80 and \$13.40, implying a 20-year

growth rate of roughly 3 percent in rural areas, but 9.3 percent in urban areas.

These figures, of course, should not be interpreted causally. That is, a highly educated worker's metropolitan status does not necessarily *cause* him to earn more than if he were situated in a smaller labor market. On the contrary, the results may reflect, at least in part, a selection mechanism by which the most productive, highly educated workers have chosen to live in cities. Still, these results seem to suggest that there are strong economic incentives for highly educated workers to reside in urban areas.

To gain a better sense of which factors (e.g., metro area size, existing human capital, education premia, industrial composition) may underlie human capital accumulation, I now turn to the analysis of a cross section of metro areas. The underlying goal is to exploit the variation exhibited across cities with respect to their education, size, and other characteristics to draw inferences about which features are most strongly associated with the growth of human capital.

## Correlates of Human Capital Accumulation: Baseline Results

As noted previously, the Census data used in this article identify more than 200 metro areas in each of the three years (1980, 1990, 2000) considered. Using this sample, I estimate the following simple regression in which the change in metro area *i*'s college fraction during decade *t*,  $\Delta Coll_{i,t}$ , is specified as

$$(1) \quad \Delta Coll_{i,t} = \mu + \delta_t + \beta X_{i,t} + \varepsilon_{i,t},$$

where  $\mu$  is a constant,  $\delta_t$  is a decade-specific fixed effect,  $X_{i,t}$  is a set of characteristics describing the metro area at the *beginning* of the decade, and  $\varepsilon_{i,t}$  is a stochastic element, assumed to be uncorrelated across metro areas but potentially correlated within them (i.e.,  $\varepsilon_{i,t}$  and  $\varepsilon_{i,s}$  may show some non-zero association). This equation is meant to be analogous to those used in empirical studies of economic growth in which a measure of growth is regressed on a set of initial characteristics (e.g., Barro, 1991, and Glaeser, Sheinkman, and Shleifer, 1995).

<sup>7</sup> The regressions also include dummies for marital status, disability status, veteran status, and foreign-born status.

<sup>8</sup> The 5 percent sample for 1990 does not report metropolitan status for all individuals in the sample. Hence, estimating the regression for this year is not possible.

<sup>9</sup> Percentages are derived from the estimates in Table 1 by exponentiating the log wage differential and subtracting 1. A 26-log-point differential between college and high school graduates in non-metro areas in 1980, for example, corresponds to roughly 30 percent.

<sup>10</sup> These estimates are based on exponentiating the coefficients in Table 1.

**Table 1**  
**Education Premia by Metropolitan Status**

Variable	1980	2000
No high school	1.84 (0.004)	1.73 (0.006)
Some high school	1.96 (0.003)	1.81 (0.005)
High school	2.09 (0.003)	1.98 (0.004)
Some college	2.15 (0.003)	2.11 (0.004)
College or more	2.35 (0.003)	2.38 (0.004)
No high school–metro	0.013 (0.004)	–0.014 (0.007)
Some high school–metro	0.031 (0.004)	0.035 (0.005)
High school–metro	0.046 (0.003)	0.053 (0.004)
Some college–metro	0.081 (0.004)	0.107 (0.004)
College–metro	0.156 (0.004)	0.215 (0.004)

NOTE: Coefficients are from regressions of log hourly wages on education indicators and their interactions with a metropolitan status dummy; 1,850,727 observations for the year 1980; 2,135,811 observations for the year 2000; standard errors appear in parentheses.

Among the characteristics considered in the vector  $X_{i,t}$  are the following: (i) an estimate of a metro area's return to a college degree,<sup>11</sup> (ii) its level of human capital (given by the fraction of college-educated workers in the labor force), (iii) its raw size (given by the logarithms of population and population density), and (iv) its broad industrial composition (measured by shares of total employment accounted for by each of 20 industries). Summary statistics for each of these regressors appear in Table 2.<sup>12</sup>

Results are given in Table 3. The first column, labeled  $I$ , reports the resulting coefficients when each covariate is entered into the regression separately. In all instances, estimation of equation (1) also includes a set of three region dummies to account for any exogenous differences in the rate

of human capital accumulation in different parts of the country.<sup>13</sup>

Based on the estimates, many of these regressors do turn out to be significantly associated with the growth of the college fraction, at least in a simple, univariate sense. Metro areas with initially larger populations, higher levels of population density, and larger fractions of workers with a bachelor's degree or more all see their college attainment rates rise by more over the following decade than smaller, less-dense, less-educated metro areas. In addition, greater fractions of employment accounted for by industries such as agriculture, mining, and manufacturing (either durable or nondurable) tend to correlate negatively with human capital accumulation, whereas a strong presence of industries such as finance, insurance, real estate, and business and repair services are positively associated with the change in the college attainment rate. Given that the former set of industries tends to employ fewer highly educated workers than the latter set of industries (see Table 4), these associations are rather intuitive. The estimated city-specific return to a college degree, while positive, is not statistically important. Greater discussion of this last regressor is provided below.

<sup>11</sup> Metro-area college degree returns are derived from city-year-specific regressions of log hourly wages on five education indicators, eight experience indicators, and dummies for marital status, disability status, veteran status, and foreign-born status. The coefficient on the college completion dummy is used to estimate the return to a college degree.

<sup>12</sup> Because they are easier to interpret, Table 2 lists summary statistics for population and population density levels rather than logarithms. In the regression analysis, I use these variables in log form, which is reasonably standard in the empirical literature on cities.

<sup>13</sup> A list of the state-level composition of the four U.S. Census regions appears in the appendix.

**Table 2**  
**Metropolitan Area Summary Statistics**

Variable	Mean	Standard deviation	Minimum	Maximum
Estimated return	2.47	0.115	2.05	2.84
Population	888,590.6	1,912,186	100,376	19,397,717
Density	578.3	1,178.4	6.01	16,258.1
College fraction	0.218	0.065	0.09	0.455
Fraction agriculture, forestry, fisheries	0.007	0.006	0.001	0.063
Fraction mining	0.006	0.016	0	0.148
Fraction construction	0.068	0.017	0.033	0.19
Fraction nondurable manufacturing	0.077	0.047	0.014	0.365
Fraction durable manufacturing	0.123	0.072	0.009	0.453
Fraction transportation	0.045	0.014	0.018	0.152
Fraction communications	0.015	0.006	0.004	0.052
Fraction utilities	0.015	0.007	0.003	0.075
Fraction wholesale trade	0.044	0.013	0.015	0.126
Fraction retail trade	0.163	0.022	0.096	0.24
Fraction finance, insurance, real estate	0.061	0.021	0.027	0.24
Fraction business and repair services	0.051	0.018	0.016	0.149
Fraction private household services	0.005	0.003	0	0.02
Fraction personal services	0.024	0.015	0.012	0.23
Fraction entertainment and recreation services	0.012	0.012	0.003	0.139
Fraction medical services	0.091	0.024	0.04	0.292
Fraction educational services	0.105	0.031	0.052	0.27
Fraction social services	0.013	0.004	0.005	0.034
Fraction other professional services	0.02	0.009	0.004	0.088
Fraction public administration	0.055	0.031	0.015	0.255

NOTE: Summary statistics are taken over 661 city-year observations.

The next two columns of results, *II* and *III*, report the coefficients from two different specifications of (1) in which various combinations of these covariates appear. The longer of these (*III*) suggests that, unlike what is reported above, very few of the initial industry shares are significantly associated with human capital accumulation. Indeed, comparing the results from columns *I* and *III*, only one industry share enters significantly in both cases: finance, insurance, real estate. Industrial composition, therefore, seems largely unimportant for explaining the growth of human capital, at least once we have conditioned on initial education, size, and returns.

Among the remaining covariates, only two show consistently positive and significant associations with human capital accumulation: log population and the initial college fraction. Both of these regressors produce significant coefficients in all three reported specifications. Log density, by contrast, becomes insignificant when industry shares are included, and the initial return to a college degree enters negatively (and significantly) in specifications *II* and *III*. This latter result may simply reflect the inverse association between various measures of urban growth (e.g., population and average earnings) and initial wages, which is a common finding in the urban economics lit-

**Table 3**  
**Human Capital Accumulation Regression Results**

Variable (initial value)	<i>I</i>	<i>II</i>	<i>III</i>
Estimated return	0.018 (0.011)	-0.02* (0.006)	-0.03* (0.014)
Log population	0.006* (0.001)	0.003* (0.001)	0.004* (0.002)
Log density	0.007* (0.001)	0.003* (0.001)	0.002 (0.002)
College fraction	0.16* (0.015)	0.12* (0.02)	0.11* (0.04)
Fraction agriculture, forestry, fisheries	-0.38* (0.13)	—	0.003 (0.17)
Fraction mining	-0.13* (0.03)	—	-0.01 (0.06)
Fraction construction	-0.002 (0.07)	—	0.18* (0.09)
Fraction nondurable manufacturing	-0.06* (0.02)	—	0.02 (0.04)
Fraction durable manufacturing	-0.03* (0.017)	—	0.03 (0.04)
Fraction transportation	0.07 (0.08)	—	0.01 (0.08)
Fraction communications	0.94* (0.19)	—	0.02 (0.23)
Fraction utilities	-0.29* (0.15)	—	-0.08 (0.15)
Fraction wholesale trade	0.01 (0.09)	—	-0.06 (0.1)
Fraction retail trade	-0.13* (0.06)	—	-0.01 (0.05)
Fraction finance, insurance, real estate	0.36* (0.06)	—	0.19* (0.07)
Fraction business and repair services	0.36* (0.11)	—	-0.15 (0.1)
Fraction private household services	-0.56 (0.42)	—	-0.3 (0.4)
Fraction personal services	0.02 (0.05)	—	0.06 (0.08)
Fraction entertainment and recreation services	0.04 (0.08)	—	0.004 (0.1)
Fraction medical services	0.03 (0.05)	—	0.03 (0.06)
Fraction educational services	-0.0005 (0.05)	—	-0.04 (0.06)
Fraction social services	0.76* (0.43)	—	0.25 (0.4)
Fraction other professional services	0.83* (0.24)	—	0.1 (0.2)
Fraction public administration	0.09* (0.03)	—	0.04 (0.05)

NOTE: The dependent variable is the change in college fraction for 1980-90 and 1990-2000. Region indicators and a dummy for the 1980-90 decade appear in all regressions. Column *I* reports coefficients from separate regressions for each regressor. Columns *II* and *III* report coefficients from regressions that include all regressors for which estimates are reported. Heteroskedasticity-consistent standard errors, adjusted for correlation within metro areas, appear in parentheses; \* denotes significance at the 10 percent level or better.

erature (e.g., Glaeser, Scheinkman, and Shleifer, 1995). Higher returns to a college degree, not surprisingly, tend to be associated with higher average wages overall in these data. As growth slows, human capital accumulation tends to slow as well.<sup>14</sup>

How significant are the estimated associations

<sup>14</sup> The positive coefficient on the initial estimated college return in specification *I* may therefore emanate from omitted-variable bias. As shown previously, returns to a college degree tend to be higher in metro areas, suggesting a positive association with population and the college attainment rate. Not including these two variables

between, on the one hand, initial log population and the college completion rate and, on the other, the subsequent change in the college completion fraction? Based on the point estimates from the longest specification in Table 3, a 1-standard-deviation increase in log population (in the cross section) corresponds to a 0.43-percentage-point rise in the college attainment rate over the next decade. A 1-standard-deviation increase in the

in specification *I* may therefore bias a truly negative coefficient on initial returns upward.

**Table 4**  
**College Attainment by Major Industry**

Industry	1980	1990	2000
Agriculture, forestry, fisheries	0.154	0.16	0.154
Mining	0.138	0.179	0.141
Construction	0.072	0.094	0.089
Nondurable manufacturing	0.112	0.153	0.196
Durable manufacturing	0.111	0.158	0.183
Transportation	0.09	0.123	0.144
Communications	0.146	0.231	0.328
Utilities	0.123	0.181	0.194
Wholesale trade	0.152	0.199	0.212
Retail trade	0.092	0.116	0.136
Finance, insurance, real estate	0.227	0.306	0.364
Business and repair services	0.2	0.255	0.33
Private household services	0.033	0.052	0.068
Personal services	0.067	0.105	0.12
Entertainment and recreation services	0.194	0.226	0.259
Medical services	0.219	0.289	0.33
Educational services	0.546	0.55	0.562
Social services	0.359	0.41	0.467
Other professional services	0.467	0.53	0.537
Public administration	0.252	0.298	0.352

NOTE: Fractions of each industry's total employment with a bachelor's degree or higher.

initial fraction of workers with a bachelor's degree or more has a somewhat larger implied association: a 0.72-percentage-point rise in the college attainment rate over the next 10 years.<sup>15</sup> Although they may seem small compared with average college completion rates near 22 percent for the metro areas in the sample, these magnitudes are far from negligible. In particular, they represent between 20 and 34 percent of the cross-sectional standard deviation of the 10-year change in the college fraction in these data, which is approximately 2.1 percentage points.

<sup>15</sup> The cross-sectional standard deviations for log population and the college completion rate are roughly 1.08 and 0.065. In terms of population levels, 1 standard deviation corresponds to roughly 680,000 residents.

## Robustness

In this section, I consider a few simple alterations to the statistical analysis to assess the robustness of the results. The first seeks to account for the influence of certain amenities (e.g., restaurants, theaters, museums) on human capital accumulation. As noted previously, Glaeser, Kolko, and Saiz (2001) have demonstrated that cities have significant consumption aspects that seem to influence the willingness of individuals to live in dense urban environments. If the highly educated have an especially strong preference for these characteristics, amenities may play an important part in human capital accumulation that the analysis above misses. Indeed, it may not be a city's population or initial level of

educational attainment that are important for explaining the growth of a city's college share, but its array of urban amenities. Population or education may simply be proxies for these types of characteristics. To explore this possibility, I consider the influence of the following eight amenities: eating and drinking establishments; movie theaters; elementary and secondary schools; live entertainment venues; museums, botanical gardens, and zoos; colleges and universities; hospitals; and commercial sports clubs (which includes professional athletics teams).<sup>16</sup> Initial values of these quantities, the first four of which are expressed in per capita terms, are added to equation (1).

Because the number of colleges and universities may not adequately capture the full extent of the college community in a metro area, I also include the total number of workers employed in these institutions. This variable should help to discern whether a metro area has, say, a particularly large university rather than a small college. In addition, although the number of elementary and secondary schools per capita is intended to serve as a rough proxy for the quality of a city's education system, it is a highly imperfect measure. As an additional proxy for school quality, I include in the regression the fraction of children between the ages of 3 and 17 who are enrolled in public school. In theory, cities with good school systems should have relatively large fractions of their school-aged children enrolled in public education. Cities with ineffective and undesirable public school systems, after all, should be characterized by higher proportions of their children attending private schools.

The second alteration takes a different approach to controlling for the influence of industrial composition. While initial shares of a metro area's employment across a broad array of sectors may offer some explanatory power with respect to human capital accumulation, how they change over time may be more relevant. That is, it may not be the initial share of employment in a city's durable manufacturing sector that affects its

college fraction, but the *change* in the fraction accounted for by that sector. Again, as demonstrated in Table 4, there are substantial differences in college attainment across the 20 industries considered. Therefore, one might expect that rising shares of employment in, say, retail trade, which employs relatively few college-educated workers, would have a negative influence on a city's overall level of education; whereas, a rise in the fraction of workers employed in educational services, which employs primarily college-educated labor, would accomplish just the opposite. To address this potential misspecification of the regression, I include contemporaneous changes in each sector's employment share in (1) and drop the initial levels.

Although this approach likely introduces a simultaneity issue into the estimation (i.e., changes in employment shares may be influenced by contemporaneous changes in the fraction of college-educated workers in the local population), it should be stressed that the objects of primary interest in this second alteration are the coefficients on log population and the initial college fraction, not those on the changes in each industry share (which, accordingly, may be biased). The idea behind this regression, quite simply, is to see whether initial size and education are still significantly correlated with subsequent changes in human capital even after removing all of the variation in human capital accumulation associated with changes in a metro area's industrial base.

Results appear in Table 5. As before, I report coefficient estimates from three different specifications to gauge the sensitivity of the findings to variations in the model. The first column, labeled *I*, reports coefficients from the regression of the change in the college attainment rate on the initial estimated return earned by college graduates, log population, log density, the initial college fraction, and initial quantities of the 10 amenities listed above.<sup>17</sup> Interestingly, five of these amenities enter significantly. Eating and drinking places per capita, live entertainment venues per capita, and numbers of colleges and universities all enter

<sup>16</sup> Many of these variables were identified by Glaeser, Kolko, and Saiz (2001) as being significantly related to population growth.

<sup>17</sup> Results were similar when the 20 initial industry shares were included. Because reporting all of these additional coefficients would have been excessive, I have omitted them from the regression.

**Table 5****Robustness Checks**

Variable	I	II	III
Initial estimated return	-0.015 (0.011)	-0.028* (0.01)	-0.027* (0.01)
Initial log population	0.005* (0.002)	0.004* (0.002)	0.004* (0.002)
Initial log density	0.002 (0.001)	0.0006 (0.002)	-0.0001 (0.001)
Initial college fraction	0.1* (0.02)	0.13* (0.02)	0.11* (0.02)
Initial eating and drinking places per capita	11.6* (3.8)	—	6.6* (4)
Initial movie theaters per capita	56.6 (72.1)	—	47.7 (72.4)
Initial live entertainment venues per capita	47.7* (23.4)	—	25.7 (24.9)
Initial elementary and secondary schools per capita	59.5 (53.3)	—	64.9 (54.5)
Initial museums, botanical gardens, zoos	-0.0003* (0.0001)	—	-0.0002 (0.0001)
Initial colleges and universities	0.0005* (0.0001)	—	0.0003* (0.0001)
Initial employment in colleges and universities	0.0008 (0.002)	—	-0.00004 (0.002)
Initial hospitals	-0.0003* (0.00007)	—	-0.0002* (0.0001)
Initial commercial sports clubs	0.0001 (0.0003)	—	0.0002 (0.0002)
Initial fraction students in public school	-0.009 (0.03)	—	-0.02 (0.03)
Δ Fraction mining	—	0.37* (0.22)	0.32 (0.23)
Δ Fraction construction	—	-0.01 (0.22)	-0.05 (0.24)
Δ Fraction nondurable manufacturing	—	0.03 (0.22)	-0.0003 (0.24)
Δ Fraction durable manufacturing	—	0.06 (0.22)	0.014 (0.23)
Δ Fraction transportation	—	0.05 (0.24)	0.08 (0.26)
Δ Fraction communications	—	-0.29 (0.29)	-0.27 (0.3)
Δ Fraction utilities	—	0.09 (0.28)	0.07 (0.29)
Δ Fraction wholesale trade	—	0.24 (0.25)	0.21 (0.26)
Δ Fraction retail trade	—	0.05 (0.22)	0.04 (0.24)
Δ Fraction finance, insurance, real estate	—	0.56* (0.25)	0.51* (0.26)
Δ Fraction business and repair services	—	0.56* (0.26)	0.53* (0.27)
Δ Fraction private household services	—	-0.04 (0.5)	0.04 (0.53)
Δ Fraction personal services	—	-0.11 (0.21)	-0.11 (0.22)
Δ Fraction entertainment and recreation services	—	0.007 (0.24)	-0.06 (0.26)
Δ Fraction medical services	—	0.28 (0.24)	0.25 (0.25)
Δ Fraction educational services	—	0.48* (0.23)	0.41 (0.25)
Δ Fraction social services	—	0.93* (0.38)	0.8* (0.4)
Δ Fraction other professional services	—	0.63* (0.36)	0.57 (0.37)
Δ Fraction public administration	—	-0.02 (0.23)	-0.06 (0.25)

NOTE: The dependent variable is the change in college fraction for 1980-90 and 1990-2000. Region indicators and a dummy for the 1980-90 decade appear in all regressions. Employment in colleges and universities is expressed in 10,000s. Heteroskedasticity-consistent standard errors, adjusted for correlation within metro areas, appear in parentheses; \* denotes significance at the 10 percent level or better.

positively; the number of museums, botanical gardens, and zoos and the number of hospitals both enter negatively.<sup>18</sup> In spite of this result, however, the coefficients on log population and the college fraction do not change appreciably from what was reported above.

The second column of results drops these 10 amenities and adds changes in 19 of the 20 industry employment shares to determine whether specifying industrial composition in 10-year differences rather than initial levels makes any difference in the remaining coefficient estimates.<sup>19</sup> Compared with the specification of industry mix in initial levels, a greater number of industries now produce significant associations, and many of these are quite reasonable, at least intuitively. An increase in the importance of finance, insurance, and real estate, as well as social and business and repair services, for example, should be associated with increases in the fraction of workers with a bachelor's degree or more. These sectors, after all, tend to employ relatively large proportions of college-educated labor. This conclusion is indeed borne out regardless of whether the 10 amenities listed above are included in the regression (column *III*) or not (column *II*).

At the same time, inclusion of changes in industrial composition has very little impact on the estimated initial population and college fraction coefficients. Both remain statistically significant, and the magnitudes are very similar to those reported in all previous specifications. Such a finding seems to reinforce the conclusion that, even after accounting for a city's industrial composition, a city's initial scale and education are strongly associated with the rate at which it accumulates highly educated workers.

Of course, characterizing the industrial composition of a metro area by using a set of 20 broad sectors is less than ideal. There is a fair amount of heterogeneity inherent in each industry; hence, this classification scheme may miss important

differences in the types of employers present in each metro area. For example, the types of employers belonging to the nondurable manufacturing sector in one city (e.g., drugs or chemicals) may be quite different from those in another (e.g., textiles or food processing). These differences may be important in explaining the growth of human capital, but would be missed by the present analysis. More seriously, these unmeasured differences may very well be directly correlated with either population or the college fraction. In such an instance, the coefficients reported thus far for these two regressors would be upwardly biased.<sup>20</sup>

I attempt to address this matter by looking, instead, at a collection of more than 200 industries, representing sectors at a mostly three-digit (standard industrial classification) level, although some two- and four-digit industries, as well as combinations of two-, three- and four-digit industries, also appear.<sup>21</sup> These are the most detailed industrial categories available in the decennial Census files.

Unfortunately, because adding more than 200 industry shares to the estimation of (1) is not practically feasible, I use the following approach: First, I create a "predicted" college attainment fraction,  $PColl_{i,t}$ , for each metro area,  $i$ , in each year  $t$ , as follows:

$$(2) \quad PColl_{i,t} = \sum_{s=1}^{N_{i,t}} Share_{s,i,t} Coll_{s,t},$$

where  $Share_{s,i,t}$  is the share of sector  $s$  in metro

<sup>18</sup> The number of hospitals may be associated with the growth in the numbers of relatively old workers who tend to possess less education than younger workers.

<sup>19</sup> Because changes in all 20 industry shares (by definition) sum to 0, I drop the change in the employment share of agriculture, forestry, and fisheries.

<sup>20</sup> For example, one city may attract human capital because it has a strong presence of nondurable manufacturing, which hires mostly highly educated workers (e.g., drugs and chemicals), whereas another may attract less human capital because it has a strong presence of nondurable manufacturing, which hires primarily less-educated workers (e.g., textiles and food processing). The presence of high- and low-human capital nondurable manufacturers will therefore be directly related to each city's initial stock of human capital, but the association between industrial composition and human capital accumulation (which is significant in this example) will be picked up by the initial stock of human capital.

<sup>21</sup> Specifically, there are 223 industries in the 1980 data, 221 in the 1990 data, and 214 in the 2000 data. These are identified using consistent codes established using the correspondence provided by the U.S. Bureau of the Census. Tobacco and crude petroleum and natural gas are examples of two-digit industries; drugs, electric light and power, and grocery stores are examples of three-digit industries; jewelry stores and retail florists are examples of four-digit industries.

**Table 6**  
**Residual College Fraction Regressions**

Variable (initial value)	I	II
Estimated return	-0.026* (0.01)	-0.024* (0.01)
Log population	0.003* (0.001)	0.004* (0.002)
Log density	0.001 (0.001)	0.0006 (0.001)
College fraction	0.08* (0.02)	0.065* (0.02)
Eating and drinking places per capita	—	5.8* (3.3)
Movie theaters per capita	—	44.3 (65.2)
Live entertainment venues per capita	—	55.7* (21.1)
Elementary and secondary schools per capita	—	38.6 (45.4)
Museums, botanical gardens, zoos	—	-0.0003* (0.0001)
Colleges and universities	—	0.0004* (0.0001)
Employment in colleges and universities	—	0.0002 (0.002)
Hospitals	—	-0.0002* (0.00006)
Commercial sports clubs	—	0.0001 (0.0002)
Fraction students in public school	—	-0.02 (0.03)

NOTE: The dependent variable is the change in the difference between a city's college fraction and its predicted college fraction based on its detailed industrial composition. Region indicators and a dummy for the 1980-90 decade appear in all regressions. Employment in colleges and universities is expressed in 10,000s. Heteroskedasticity-consistent standard errors, adjusted for correlation within metro areas, appear in parentheses; \* denotes significance at the 10 percent level or better.

area  $i$ 's total employment in year  $t$ ,  $Coll_{s,t}$  is the college completion fraction for sector  $s$  in year  $t$  (calculated using aggregate data for the United States), and  $N_{i,t}$  is the number of sectors in metro area  $i$  in year  $t$ . Second, I compute a "residual" college fraction given by  $(Coll_{i,t} - PColl_{i,t})$ , which measures the difference between a city's actual college-completion fraction and the fraction that would result if its industries resembled the national average. I interpret this difference as the part of a city's college-attainment fraction that is not explained by its detailed industry composition. I then consider regressions of the form

$$(3) \quad \Delta(Coll_{i,t} - PColl_{i,t}) = \mu + \delta_t + \beta X_{i,t} + \varepsilon_{i,t},$$

where two specifications of the regressors  $X_{i,t}$  are considered: One controls for the estimated college return, log population, log density, and the college fraction, all in initial levels; the other further adds initial values of the 10 amenities discussed above. The resulting estimates appear in Table 6.

In general, they demonstrate very little change

from what has already been reported. Among the amenities, the same five variables (eating and drinking places per capita; live entertainment venues per capita; numbers of museums, botanical gardens, and zoos; numbers of colleges and universities; and numbers of hospitals) all enter significantly and with the same signs as before. Additionally, the initial college-return produces a significantly negative coefficient, while the logarithm of population and the initial fraction of college-educated workers in total employment generate significantly positive coefficients.

With these latter two regressors, it is worth noting that the coefficients are now somewhat smaller than what is reported in Tables 3 and 5. For example, in Table 6, log population produces coefficients between 0.003 and 0.004 rather than between 0.003 and 0.006 previously, whereas the initial college completion rate generates a coefficient ranging from 0.065 to 0.08 rather than from 0.1 to 0.16. These decreases are consistent with the idea mentioned previously that using

**Table 7**  
**Growth Regressions**

Dependent variable	Specification	Initial college fraction	Initial log population	Initial average log hourly wage
Population growth	<i>I</i>	0.28* (0.13)	—	—
	<i>II</i>	—	0.011* (0.005)	—
	<i>III</i>	0.41* (0.14)	0.02* (0.006)	-0.33* (0.08)
Average hourly wage growth	<i>I</i>	0.26* (0.04)	—	—
	<i>II</i>	—	—	-0.11* (0.035)
	<i>III</i>	0.42* (0.05)	0.02* (0.003)	-0.39* (0.05)

NOTE: Regressions of metro area-level population growth and average hourly wage growth on initial values of the college fraction, log population, and average log hourly wages. Region indicators and a dummy for the 1980-90 decade appear in all regressions. Heteroskedasticity-consistent standard errors, adjusted for correlation within metro areas, appear in parentheses; \* denotes significance at the 10 percent level or better.

20 broad industry shares leads to upwardly biased coefficients on the initial college fraction and log population. Still, the evidence is remarkably consistent with respect to the influence of these two variables. Regardless of how the statistical model is specified, initial population and education are significant predictors of human capital accumulation.

### **Human Capital, Growth, and Divergence**

The finding that more-populous and -educated cities tend to experience the largest increases in human capital has an intriguing implication with respect to the geographic distributions of population and college-educated labor. Specifically, it suggests that the distributions of these two quantities should have been characterized by increasing concentration over the 1980-2000 period. Human capital accumulation, after all, tends to be faster in cities with larger initial fractions of highly educated workers. This mechanism should then lead to a growing gap between the education levels across cities over time as the top end of the distribution pulls away (or “diverges”) from the bottom. Because previous work has shown that more-educated cities also tend to see faster population growth (e.g., Glaeser, Scheinkman, Shleifer, 1995), I arrive at a similar implication with respect to the distribution of population. This section exam-

ines whether there has been this type of “divergence” in the distribution of these two quantities.

Before doing so, I attempt to establish some basic results relating the growth of two quantities—population and average hourly wages—to education. While the former is of greater interest in this particular exercise, the latter more closely resembles the object of interest in most studies of economic growth (i.e., per capita income). Results from the regression of each quantity’s 10-year growth rate on the initial level of human capital appear in the specifications labeled *I* in Table 7.<sup>22</sup> Not surprisingly, each shows a significantly positive association with initial education. Here, the magnitudes indicate that a 1-standard-deviation (i.e., a 6.5-percentage-point) increase in a city’s college attainment rate tends to be accompanied by a 1.8-percentage-point rise in its rate of population growth and a 1.7-percentage-point rise in its rate of average wage growth over the next 10 years. These figures represent, respectively, 16 and 20 percent of the cross-sectional standard deviations in these two growth series. These associations, therefore, seem to be both statistically and economically important.

To explore whether there has been divergence across city-level human capital, population, and

<sup>22</sup> As with all of the other regressions, these include three region dummies and an indicator for the 1980-90 decade.

average wages, I consider two approaches. The first looks for so-called  $\beta$ -convergence, the test for which involves a simple regression of the growth of a quantity on its initial level.<sup>23</sup> A negative coefficient on the initial level of a variable would indicate a tendency for that quantity to converge to a common level across metro areas. After all, a negative coefficient would indicate that cities with low levels of human capital, for example, would experience faster human capital growth than cities with high levels. This process should generate a less-concentrated distribution of human capital over time as the bottom of the distribution catches up with the top. The second approach looks for  $\sigma$ -convergence, which is based on how the cross-sectional dispersion of a particular quantity changes over time. Decreasing dispersion (i.e., falling concentration) would be indicative of  $\sigma$ -convergence.<sup>24</sup>

One common criticism of these statistical approaches, particularly tests for  $\beta$ -convergence, pertains to the appropriateness of pooling a set of extremely heterogeneous economies in the same regression (see Durlauf and Quah, 1999). While this point is certainly valid when considering studies of countries, which tend to vary substantially in terms of various fundamental characteristics including how their economies function (e.g., Japan and Nigeria), it is less likely to be a significant issue when comparing the experiences of metro areas within the same country (e.g., Seattle and Atlanta).

The  $\beta$ -convergence results for metro-area college attainment are already well-established in the findings shown thus far. The strong positive association between the initial level of a city's college fraction and its subsequent change over the next decade indicates divergence in this variable. Results for the logarithm of population and the average log hourly wage appear in the specifications labeled *II* in Table 7.

<sup>23</sup> Again, all regressions also include three region dummies and a time effect to pick up differences in growth across decades. The  $\beta$  in  $\beta$ -convergence refers to the coefficient on the initial level of a variable in a growth regression.

<sup>24</sup> The  $\sigma$  in  $\sigma$ -convergence refers to the standard deviation. Barro and Sala-i-Martin (1995) provide an overview of the statistical techniques commonly used in studies of convergence/divergence.

The population series also shows divergence which, intuitively, is precisely what one would expect in light of the results shown to this point. Larger populations tend to be associated with more rapid human capital accumulation, which raises education levels. This, in turn, leads to faster population growth. Hence, one would expect to see a positive association between initial population and its subsequent rate of growth. Interestingly, however, the positive association between initial population and its subsequent growth also holds after conditioning on the initial college fraction and the initial average log hourly wage. This result is reported in specification *III*. The direct association between population and population growth, therefore, does not seem to be driven entirely by education. There is some aspect of metro area size that, independent of education, draws additional population.

Average hourly wages, by contrast, show evidence of convergence rather than divergence. That is, higher average wages tend to be followed by slower rates of wage growth over the next decade. This finding, too, is sensible given the evidence already presented. Recall that higher wages tend to be accompanied by slower subsequent human capital accumulation. The significantly negative coefficients on the initial college return in the regression results presented above demonstrate this point clearly. Slower human capital accumulation, then, implies slower growth of average hourly wages. Thus, one would expect to see a negative association between initial average wages and future wage growth. This relationship turns out to hold whether initial education and log population are accounted for or not (compare specifications *II* and *III*).

To look at  $\sigma$ -convergence, I need a measure that characterizes the degree of spread in the distributions of human capital, log population, and the average log hourly wage.<sup>25</sup> In an effort to keep the analysis broad, I consider several possible

<sup>25</sup> For this exercise, I use population and average wages in logarithmic form because the distributions of their levels will tend to show increasing dispersion even if growth is unrelated to the initial level. For example, the gap between the populations of two cities, one with population of 100, the other with a population of 1000, will grow wider if both cities grow by the same percentage (and possibly if the smaller city grows by a larger percentage).

**Table 8**  
**Features of the Education, Population, and Average Hourly Wage Distributions**

Variable	Statistic	1980	1990	2000	
College fraction	Mean	0.178	0.219	0.253	
	10th percentile	0.122	0.148	0.17	
	25th percentile	0.15	0.18	0.206	
	50th percentile	0.175	0.216	0.248	
	75th percentile	0.203	0.25	0.284	
	90th percentile	0.238	0.296	0.339	
	Standard deviation	0.043	0.054	0.064	
	90-10 difference	0.116	0.148	0.168	
	90-50 difference	0.063	0.08	0.091	
	50-10 difference	0.053	0.068	0.077	
	75-25 difference	0.053	0.069	0.078	
	Log population	Mean	12.96	13.06	13.19
		10th percentile	11.76	11.79	11.91
25th percentile		12.07	12.17	12.36	
50th percentile		12.77	12.88	13.01	
75th percentile		13.61	13.71	13.9	
90th percentile		14.37	14.62	14.76	
Standard deviation		1.057	1.075	1.092	
90-10 difference		2.61	2.83	2.86	
90-50 difference		1.6	1.74	1.75	
50-10 difference		1	1.09	1.1	
75-25 difference		1.54	1.54	1.54	
Average log hourly wage		Mean	2.46	2.45	2.5
		10th percentile	2.33	2.32	2.43
	25th percentile	2.39	2.38	2.5	
	50th percentile	2.46	2.45	2.55	
	75th percentile	2.51	2.51	2.61	
	90th percentile	2.58	2.58	2.68	
	Standard deviation	0.098	0.109	0.105	
	90-10 difference	0.24	0.26	0.25	
	90-50 difference	0.12	0.14	0.13	
	50-10 difference	0.13	0.13	0.12	
	75-25 difference	0.12	0.13	0.11	

NOTE: Statistics are based on 188 metro areas for the college fraction and average log hourly wage and 187 metro areas for log population.

## Wheeler

measures: the standard deviation and a host of inter-quantile differences (e.g., the difference between the 90th percentile and the 10th). One important consideration in looking at these distributional features is maintaining a consistent sample of metro areas. The sample of metro areas identified by the Census does change from one year to the next. As a consequence, there may be changes in the degree of spread in the distribution of these variables that stem from changes in the composition of the sample rather than an actual convergence or divergence mechanism. In computing these distributional features, then, I confine the sample to those 188 metro areas that appear in all three years.

The resulting estimates appear in Table 8. Looking at the distribution of college attainment rates, it is evident that, although there has been an increase in the fraction of workers with a bachelor's degree or more at all points of the distribution, that increase has been larger at the top than at the bottom. The 90th percentile, for example, rose by more than 10 percentage points between 1980 and 2000, increasing from 0.238 to 0.339. The corresponding increases for the median and 10th percentiles over this period were 7.3 and 4.8 percentage points. Accordingly, each of the four listed percentile gaps (90-10, 90-50, 50-10, 75-25) grew wider over time. Rising dispersion can also be inferred from the evolution of the standard deviation, which started at 0.043 in 1980, rose to 0.054 in 1990, and stood at 0.064 by 2000. Evidently, human capital became more unevenly distributed during this time frame.<sup>26</sup>

The logarithm of population, the distributional features of which appear just below the human capital results in Table 8, reveals a similar trend. On average, metro areas in the United States experienced population gains between 1980 and 2000, and these gains were registered at all five quantiles of the distribution. Again, however, the gains tended to be somewhat larger at the top of the distribution than at the bottom. With the exception of the inter-quartile difference (75-25), which did not change between 1980 and 2000,

all other quantile differentials increased in both decades. Increasing dispersion in the logarithm of population also shows up in the standard deviations, which increased from 1.057 in 1980 to 1.075 in 1990 to 1.092 in 2000. Log population, therefore, also shows evidence of both types of divergence.

These results are particularly striking because they stand somewhat at odds with what conventional economic analysis might suggest. Indeed, as cities grow in population, they tend to become more congested, which, in turn, raises costs (financial and otherwise) to both workers and employers. These “agglomeration diseconomies” should, therefore, work to slow subsequent rates of population growth as firms and workers seek less-congested labor markets. Similarly, as the fraction of workers with a bachelor's degree rises, the relative return received by college-educated workers (all else equal) should decline because the supply of such workers has risen relative to demand. This is a standard diminishing marginal productivity argument whereby the return received by a factor of production (e.g., college-educated labor) declines as it is used more intensively relative to all other inputs. A lower return paid to college-educated labor, of course, should reduce the rate at which workers with a bachelor's degree move into an area. Empirically, however, there seems to be little support for these theoretical ideas in the data.

Recall that average hourly wages show a very different pattern. Regressions of wage growth on initial wage levels reveal a significantly negative relationship between the two. One might expect, therefore, to see a decrease in the degree of dispersion in the distribution of city-level average log wages. The estimated dispersion measures in Table 8, however, show only a decrease between 1990 and 2000, when the standard deviation and all four quantile differences narrowed. During the 1980s, all but the 50-10 difference increased. These results demonstrate an important difference between  $\beta$ - and  $\sigma$ -convergence. Although a negative association between the initial level of a variable and its subsequent growth rate may certainly reduce the degree of variance in a distribution, it may also increase it. Durlauf and Quah

<sup>26</sup> Moretti (2004) documents a similar rise in the degree of human capital “inequality” across U.S. metro areas.

(1999), for example, show how  $\beta$ -convergence may generate a wider distribution if economies with low levels of a variable overshoot economies with high levels. Distributional dynamics of this sort may help to explain these results.

Another possible explanation may relate to the influence of population and the college fraction, which, as shown in Table 7, tend to be positively related to wage growth over the next decade. As these two variables have diverged, they may have led to a divergence in wage levels during the 1980s if their influence outweighed the natural tendency for wage levels to converge. Assuming that this natural tendency was stronger during the 1990s, of course, the cross-sectional dispersion in average log wages would have declined.<sup>27</sup>

## CONCLUDING DISCUSSION

This paper has explored the issue of human capital accumulation across a set of U.S. metro areas. Among the more prominent findings is that cities with larger populations and larger fractions of workers with college degrees tend to see faster growth in their stocks of human capital. Because human capital also tends to be positively associated with population growth, this process has led to a divergence of both human capital and population in the United States between 1980 and 2000. Hence, the largest and most-educated cities in the country have tended to accumulate population and human capital faster than smaller and less-educated cities.

The divergence of human capital and population has not, interestingly, generated much divergence with respect to average wage levels across cities. Although the amount of dispersion in the distribution of metro area–level average wages did grow larger between 1980 and 2000, this growth occurred during the decade of the 1980s. Dispersion in wage levels actually declined somewhat between 1990 and 2000. This result could

be related to the mechanism described above in which the increasing concentration of workers with college degrees may depress the returns they receive.

If true, however, why would college-educated workers continue to flock to labor markets with large populations and stocks of highly educated workers? Glaeser (1999) suggests that workers with a college degree may seek to surround themselves with other college graduates because they are able to learn from one another. Highly educated workers, according to this line of reasoning, are especially committed to the acquisition of productive skills. Since previous work suggests that there may be productive externalities associated with the presence of college-educated individuals (e.g., Moretti, 2004), the positive association between initial college attainment rates and subsequent changes in these rates may reflect the desire of highly educated workers to reside in environments that facilitate learning.

Peri (2002) echoes this view, suggesting that, if skill acquisition is an important reason for the concentration of human capital in cities, we should expect to see large numbers of “young” college-educated workers in cities. Young workers, after all, are more likely than their older counterparts to seek learning opportunities because they are in the early stages of their careers and, therefore, know relatively little. The evidence he reports is certainly consistent with this idea. Between 1970 and 1990, the ratio of college-educated workers with fewer than 20 years of work experience to those with more than 20 years rose from 1.5 to 2.12 within the metro areas of the United States.

Populous cities may also help facilitate the job search process for highly educated married couples. Costa and Kahn (2000) suggest that “power couples” (i.e., those in which both partners have a bachelor’s degree or more) have increasingly moved into large metro areas over the past several decades because cities are more likely to offer job opportunities for both spouses. Large cities, therefore, may provide a solution to the occupational co-location problem.

An additional possibility that deserves to be mentioned involves the amenity value of college-educated workers themselves. That is, while the

<sup>27</sup> A similar argument relating to the relative strengths of the college attainment rate, population, and average wage levels could be made in explaining the divergence patterns of human capital and log population. In those cases, evidently, the mechanisms leading to divergence were stronger than any effects that wage levels might have had.

## Wheeler

college-educated may be enticed to locate in cities with a large presence of certain types of establishments (say, eating and drinking places), they might also want to be around other college-educated workers because they desire homogeneity in their social interactions. So, even though concentrations of highly educated workers may be associated with diminishing returns and lower earnings (at least, all else equal), college graduates may still want to surround themselves with other highly educated workers because they find them to be desirable neighbors. Of course, a strong presence of college-educated workers may also be associated with characteristics that have not been accounted for directly here (e.g., low crime, greater civic engagement, good schools), but which are especially desirable to these types of workers.

Whatever the reason, divergence in the distribution of human capital may, over longer time horizons, begin to lead to divergence in earnings and productivity across cities and regions. Although there is only limited evidence of that having occurred between 1980 and 2000, it may occur to a greater extent in future decades. In particular, if technologies respond to the supply of skills (e.g., as suggested by Acemoglu, 1998), cities with large stocks of college-educated labor may increasingly use technologies that complement these types of workers. This trend may further reinforce the divergence of human capital by encouraging highly educated workers to congregate in the most educated cities as well as lead to greater productivity differentials between cities with small stocks of human capital and those with large stocks.

## REFERENCES

- Acemoglu, Daron. "Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality." *Quarterly Journal of Economics*, November 1998, 113(4), pp. 1055-89.
- Barro, Robert J. "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics*, May 1991, 106(2), pp. 407-43.
- Barro, Robert J. and Sala-i-Martin, Xavier. "Convergence." *Journal of Political Economy*, April 1992, 100(2), pp. 223-51.
- Barro, Robert J. and Sala-i-Martin, Xavier. *Economic Growth*. New York: McGraw-Hill, 1995.
- Black, Duncan and Henderson, Vernon. "A Theory of Urban Growth." *Journal of Political Economy*, April 1999, 107(2), pp. 252-84.
- Card, David. "The Causal Effect of Education on Earnings," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*. Volume 3A. Amsterdam: Elsevier, 1999, pp. 1801-63.
- Costa, Dora L. and Kahn, Matthew E. "Power Couples: Changes in the Locational Choice of the College Educated, 1940-1990." *Quarterly Journal of Economics*, November 2000, 115(4), pp. 1287-315.
- Dee, Thomas S. "Are There Civic Returns to Education?" *Journal of Public Economics*, August 2004, 88(9-10), pp. 1697-720.
- Durlauf, Steven N. and Quah, Danny T. "The New Empirics of Economic Growth," in John B. Taylor and Michael Woodford, eds., *Handbook of Macroeconomics*. Volume 1A. Amsterdam: Elsevier, 1999, pp. 234-308.
- Glaeser, Edward L. "Learning in Cities." *Journal of Urban Economics*, September 1999, 46(2), pp. 254-77.
- Glaeser, Edward L.; Kolko, Jed and Saiz, Albert. "Consumer City." *Journal of Economic Geography*, January 2001, 1(1), pp. 27-50.
- Glaeser, Edward L. and Saiz, Albert. "The Rise of the Skilled City." Discussion Paper No. 2025, Harvard Institute of Economic Research, 2003.
- Glaeser, Edward L. and Saks, Raven E. "Corruption in America." Discussion Paper No. 2043, Harvard Institute of Economic Research, 2004.
- Glaeser, Edward L.; Scheinkman, José A. and Shleifer, Andrei. "Economic Growth in a Cross-Section of Cities." *Journal of Monetary Economics*, December 1995, 36(1), pp. 117-43.

- Lochner, Lance and Moretti, Enrico. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review*, March 2004, 94(1), pp. 155-89.
- Milligan, Kevin; Moretti, Enrico and Oreopoulos, Philip. "Does Education Improve Citizenship? Evidence from the United States and the United Kingdom." *Journal of Public Economics*, August 2004, 88(9-10), pp. 1667-95.
- Moretti, Enrico. "Human Capital Externalities in Cities," in J. Vernon Henderson and Jacques-Francois Thisse, eds., *Handbook of Regional and Urban Economics*. Volume 4. New York: Elsevier, 2004, pp. 2243-91.
- Park, Jin Heum. "Estimation of Sheepskin Effects and Returns to Schooling Using the Old and the New CPS Measures of Educational Attainment." Princeton Industrial Relations Section Working Paper No. 338, Princeton University, 1994.
- Peri, Giovanni. "Young Workers, Learning, and Agglomerations." *Journal of Urban Economics*, November 2002, 52(3), pp. 582-607.
- Ruggles, Steven; Sobek, Matthew; Alexander, Trent; Fitch, Catherine A.; Goeken, Ronald; Hall, Patricia Kelly; King, Miriam and Ronnander, Chad. *Integrated Public Use Microdata Series: Version 3.0* [machine-readable database]. Minneapolis, MN: Minnesota Population Center, 2004.
- Simon, Curtis J. and Nardinelli, Clark. "Human Capital and the Rise of American Cities, 1900-1990." *Regional Science and Urban Economics*, January 2002, 32(1), pp. 59-96.
- U.S. Bureau of the Census. *USA Counties 1998*, on CD-ROM [machine readable data file]. Washington, DC: U.S. Bureau of the Census, 1999.

---

## APPENDIX

### *Census Data*

The data are taken from the 1980, 1990, and 2000 5 percent samples of the Integrated Public Use Microdata Series (IPUMS) described by Ruggles et al. (2004). Specifically, I use the 1980 5 percent state sample, the 1990 5 percent state sample, and the 2000 5 percent sample. These files have roughly 11.3 million, 12.5 million, and 14 million observations, respectively.

To calculate educational attainment distributions among each metropolitan area's labor force, I focus on the working age population (i.e., those who are between the ages of 18 and 65) who report positive wage and salary earnings and are not in school at the time the Census was taken. In estimating the returns to various levels of formal schooling, I further limit the analysis to white males who earn between \$1 and \$250 per hour. This trimming procedure is designed to eliminate the influence of outlier observations, which occasionally appear due to the computation of hourly earnings as the ratio of annual wage and salary earnings to the product of weeks worked and usual hours per week. All dollar figures are converted to real terms (year 2000 dollars) using the personal consumption expenditure chain type price index.

Potential experience is calculated as the maximum of (age minus years of education minus 6) and 0. Because years of education is not reported for all individuals in the 1990 and 2000 Census, where educational attainment is sometimes reported as a range, I have imputed years of schooling completed using Table 5 of Park (1994).

As noted in the text, metro areas are defined as metropolitan statistical areas (MSAs), New England county metropolitan areas (NECMAs), or consolidated metropolitan statistical areas (CMSAs) if an MSA or NECMA belongs to a CMSA. Although somewhat large when considering local labor markets

## **Wheeler**

(e.g., New York– northern New Jersey–Long Island), CMSAs greatly facilitate the construction of geographic areas that have reasonably consistent geographic boundaries. Due to changes in geographic definitions, residents of one MSA within a CMSA in a particular year are sometimes categorized as residing in a different MSA (within the same CMSA) in another year. Aggregating to the CMSA level eliminates any problems arising from this type of definitional change.

Across the 210 metro areas identified in the 1980 Census, the average number of individual level observations used to construct the college attainment and industry share statistics is 14,484.1 (minimum = 1,557, maximum = 315,128). Among the 206 metro areas identified in the 1990 data, the average is 15,863.4 (minimum = 1,426, maximum = 329,632). In the 245 metro areas from the 2000 Census, the mean is 16,526.4 (minimum = 1,426, maximum = 371,278). When confining the sample to white males only for the college return regressions, the mean numbers of observations (minimum, maximum) per metro area are 6,928.1 (666, 144,886) for 1980, 7,275.7 (670, 144,698) for 1990, and 6,600.7 (516, 134,898) for 2000.

A complete list of the detailed industries appears in the IPUMS documentation at [www.ipums.org](http://www.ipums.org). These can be found in the link to the industry codes for 1980. The 1990 and 2000 codes are translated into equivalent 1980 codes using the programs that accompany this paper.

### ***Additional Data Details***

Metro-area population density is calculated as a weighted average of county-level densities, where the weights are given by population shares. This measure is intended to give a better sense of the average density per square mile faced by a typical city dweller than average density (metro area population to metro area land area), which may be misleading, particularly among cities in the western United States, which encompass extremely large, but sparsely populated counties.

### ***U.S. Census Regions***

Midwest: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin

Northeast: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont

South: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia

West: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming