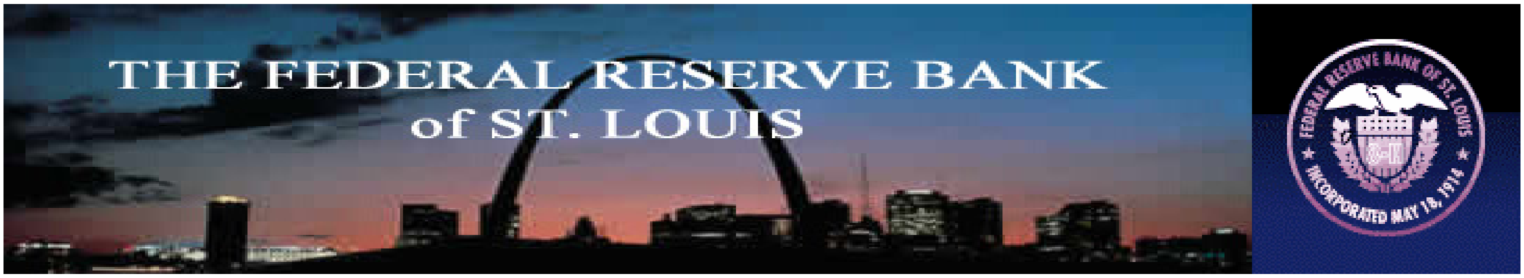

WORKING PAPER SERIES

Do Localization Economies Derive from Human Capital Externalities?

Christopher H. Wheeler

Working Paper 2005-015A
http://research.stlouisfed.org/wp/2005/2005-015.pdf

February 2005

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
411 Locust Street
St. Louis, MO 63102

Do Localization Economies Derive from Human Capital Externalities?

Christopher H. Wheeler*
Research Division
Federal Reserve Bank of St. Louis
P.O. Box 442
St. Louis, MO 63166
christopher.h.wheeler@stls.frb.org

February 15, 2005

Abstract

One of the most robust findings emerging from studies of industrial agglomeration is the rise in productivity that tends to accompany it. What most studies have not addressed, however, is the potential role played by human capital externalities in driving this relationship. This paper seeks to do so using data from the 1980, 1990, and 2000 US Census covering a collection of 77 (primarily) 3-digit manufacturing industries across a sample of more than 200 metropolitan areas. The analysis generates two primary results. First, a variety of education- and experience-based measures of average human capital rise significantly as an industry's employment in a metropolitan area increases. Hence, clusters of industry do tend to be characterized by larger stocks of human capital. However, second, even after accounting for the level of human capital in a worker's own industry, the overall size of the industry remains strongly associated with wages. Such results suggest that localization economies are largely not the product of knowledge spillovers.

JEL Classification: R11
Keywords: Human Capital Externalities, Agglomeration Economies, Industrial Localization

*The views expressed herein are those of the author and do not represent the official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

1 Introduction

Productivity gains tied to the geographic concentration of industry (i.e. 'localization') are among the most robust empirical findings in the urban economics literature. Regardless of the productivity measure considered (e.g. output per worker, total factor productivity, wages) or the level of aggregation on which the analysis is based (e.g. plants, workers, aggregate city- or state-industries), studies uniformly find that productivity rises significantly as an industry's presence within a local labor market grows (e.g. Carlino (1979), Henderson (1986, 2003), Wheeler (2004)).[1]

Traditionally, this empirical regularity has been interpreted as evidence of Marshallian externalities: that is, external productivity shifts that Alfred Marshall (1920) suggested were the product of (*i*) the spillover of industry-specific knowledge across producers, (*ii*) the creation of a specialized input-producing sector, and (*iii*) a more efficient firm-worker matching mechanism. Through each of these channels, localization is hypothesized to offer a residual boost to production, increasing the output that an individual producer generates with a given set of inputs.

A related literature, grounded more in the theory of economic growth than urban economics, holds that human capital also has important external effects on productivity. Following the models of Romer (1986) and Lucas (1988), a large literature has emerged over the past decade connecting long run economic growth to the positive external effects inherent in the accumulation of knowledge and skills. A larger base of knowledge, after all, should

[1]The magnitudes of these associations do vary somewhat depending on the productivity and localization measures considered, yet most tend to be sizable. Henderson (1986), for example, estimates output-employment elasticities in excess of 0.1 - that is, a 10 percent increase in local industry employment corresponds to a 1 percent rise in output, all else equal - using aggregate city-industry manufacturing data. Wheeler (2004) estimates wage-employment elasticities between 0.02 and 0.03 using worker-level data from US metropolitan areas.

make workers in an economy more productive by exposing them to greater quantities of information which, presumably, allows them to learn more quickly.

Of course, due to the local nature of human interactions, these types of spillovers are likely to be confined to relatively small geographic areas.[2] Hence, many empirical studies of this issue over the past decade have examined the relationship between productivity (usually quantified by wages) and human capital using a variety of local labor markets (primarily cities) as economies.[3] The findings, to a large degree, suggest that human capital externalities may be quite sizable. Rauch (1993), for instance, estimates that, after conditioning on a host of personal and city-specific observables, an additional year of schooling among the residents of a metropolitan area is associated with a 3 percent rise in average hourly wages.[4] Moretti (2004b) finds that, after accounting for a plant's own inputs, an 1-percentage point increase in the share of workers with a college degree in the local market (outside the plant's own industry) correlates with a 0.8 percent rise, roughly, in its output. While there remains some disagreement about the true significance of these estimates, virtually all studies of this issue using local labor markets in the US have found some evidence of a strong positive association between aggregate human capital and productivity.

What few studies have considered, however, is whether these two empirical regularities are in some way related. In particular, little work on industrial concentration has explored the possibility that localization economies may be the reflection of intra-industry human capital spillovers. Such neglect is surprising in light of the following two well-established

[2] Jaffe et al. (1993) have provided evidence on this matter with respect to patent citations.

[3] Among the more prominent examples are Rauch (1993), Acemoglu and Angrist (2000), Moretti (2004a, 2004b), and Ciccone and Peri (2004). With the exception of Moretti (2004b), which explores output-based productivity measures using plant-level data, all examine the relationship between wages and human capital. Also, other than Acemoglu and Angrist (2000), who look at US states, all others examine externalities in metropolitan areas.

[4] Moretti (2004a) reports similar figures using data from both the US Census and National Longitudinal Survey of Youth.

results.

First, localization - quantified either by total industry-employment within a local market or the degree to which an industry is over-represented in an area's total employment when compared to the national average - is strongly associated with average establishment size. Evidence reported by Kim (1995), Holmes and Stevens (2002), and Wheeler (2004) indicates that industrial clusters tend to be populated by large numbers of producers who employ, on average, large numbers of workers. Wheeler's (2004) estimates from two- and three-digit US manufacturing, for instance, suggest elasticities of average employment per plant with respect to industry employment between 0.6 and 0.9.

Second, large establishments tend to employ more educated workers than small establishments. Idson (1989), Dunne (1994), Doms et al. (1997), Oi and Idson (1999), Troske (1999), and Moretti (2004b) (among many others) all report evidence that large plants hire more skilled (at least, more educated) employees than small plants. For example, looking at data on non-union workers employed in the private, non-agricultural sector of the US, Idson (1989) reports a mean of 11.43 years of schooling among plants with 1 to 24 employees. The corresponding averages for plants with 25 to 99, 100 to 999, and 1000 or more employees are 12.46, 13.59, and 14.07 years.

One of the reasons we observe a positive connection between industry localization and wages, therefore, may be that clusters of industry are characterized by relatively large stocks of human capital. This paper seeks to evaluate this conjecture by exploring two rather simple, yet fundamental issues: first, whether there is any significant relationship between an industry's local market scale and its level of human capital; and second, whether the estimated association between labor earnings and industrial concentration can be explained by human capital externalities. As far as I am aware, this paper is the first to do so.

Using micro-data from the 1980, 1990, and 2000 US Census on workers employed in 77 (mostly three-digit) manufacturing industries across a sample more than 200 metropolitan areas, I find two primary results. First, education- and experience-based measures of average human capital do indeed rise with city-industry employment.[5] While not overwhelmingly large, the estimated magnitudes of the associations are far from trivial: a 1 standard deviation increase in city-industry employment, for example, corresponds to a, roughly, 1 percentage point increase in the fraction of workers with a college degree and a 0.7-year increase in average work experience. Given Moretti's (2004a) estimates associating a 1 percentage point increase in a worker's city-level college fraction with a 1 to 1.5 percent increase, approximately, in his or her wages, this result suggests that human capital externalities may very well account for a significant part of the localization-wage correlation.[6]

The data, however, only show limited support for this idea because, second, the estimated localization elasticities from standard hedonic wage regressions change only slightly once I condition on industry-specific human capital. Without controlling for human capital, localization 'effects' average approximately 0.028, which is similar to what previous research has documented. Conditional on human capital, the average elasticity only drops to roughly 0.024. Although certainly not negligible, such a change is small - on the order of 14 percent of the unconditional estimate - which suggests that localization economies, in large part, do *not* reflect human capital externalities. Mechanisms operating independently of human capital appear to account for the majority of the localization phenomenon.

The remainder of the paper is organized as follows. The next section provides a brief discussion of the data. Section 3 then presents the results. Section 4 concludes with a short discussion of what this evidence may imply with respect to theories of localization.

[5] Although metropolitan areas are the geographic unit of analysis in this paper, I use the terms 'city' and 'metropolitan area' interchangeably for expositional purposes.

[6] Based on Wheeler's (2004) estimates, for instance, 1.5 percent is approximately one quarter to one third of the implied localization association.

2 Data

Individual-level data are taken from the 1980, 1990, and 2000 US Census as prepared by the Integrated Public Use Microdata Series (IPUMS).[7] Because the vast majority of previous work on industrial agglomeration has focused on manufacturing, I restrict the analysis to workers employed in this sector.

For the sake of the wage analysis conducted below, I utilize the IPUMS 1 percent samples for each year. From these, I limit the observations to white males between the ages of 18 and 65 who reported working at least 30 hours per week for at least 14 weeks during the previous year and who were not in school at the time of the survey.[8] Doing so confines the analysis to individuals with a relatively strong attachment to the labor force (i.e. their primary activity is work) and eliminates the need to control for earnings differentials based on race and gender. After further eliminating individuals for whom some of the basic covariates used below were not identified (e.g. metropolitan area of residence), I arrive at a sample of 176084 observations over the three years.

Local labor markets are taken to be metropolitan areas, 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, the use of CMSAs facilitates the creation of metropolitan areas with consistent definitions over time. Of the 275 such areas that exist in the US (using 1995 definitions), 219 distinct metropolitan areas appear in the sample used in the analysis below.

Industries are defined by the Census three-digit code. For the most part, these correspond to three-digit Standard Industrial Classification (SIC) industries, although some

[7]For details, see Ruggles and Sobek et al. (2003).

[8]While I use the 5 percent samples to calculate human capital in city-industries, the 5 percent samples are extremely large and, thus, make the estimation of the wage regressions (described below) very difficult.

represent two-digit, four-digit, or combinations of three- and four-digit sectors. For example, Dairy Products (Census code 101) and Drugs (181) are also three-digit SIC industries. However, Tobacco products (130) is a two-digit industry; Pulp, Paper, and Paperboard Mills (160) represents a group of three-digit industries; Primary Aluminum Industries (272) and Electronic Computing Equipment (322) are collections of four-digit sectors. In all, a total of 77 manufacturing industries are represented in the final sample.

Because the calculation of human capital within reasonably detailed city-industries requires large numbers of observations, I use the 5 percent IPUMS samples from each year for this purpose. From them, I compute four measures of city-industry human capital: mean years of education, mean years of experience, the fraction of college educated workers in total employment, and the fraction of total hours worked accounted for by college educated workers.[9] To maximize the number of observations used to compute these quantities, all individuals with positive wage and salary earnings who report an industry of employment are used in the calculations. Additional details about these data appear in the Appendix.

Data on city-industry employment is calculated from three County Business Patterns (CBP) files which cover the years 1980, 1990, and 1997. Because the CBP data were compiled according to the SIC system prior to 1998, but the North American Industry Classification System (NAICS) thereafter, I use the 1997 CBP data instead of the 2000 data for the sake of consistently matching city-industries across the two data sets.[10] I assume that the 1997 figures provide a reasonable approximation to the 2000 data.[11] Metropolitan

[9]The education measures, not surprisingly, tend to be positively correlated with one another: 0.99 for the employment and hours fractions, 0.66 for the employment fraction and mean education years, 0.65 for the hours fraction and mean education years. Mean years of experience, however, varies inversely with each of these: -0.2 for mean education years, -0.07 for the two college fractions. This result very likely reflects the fact that older workers in the sample tend to possess lower levels of education.

[10]The SIC and NAICS are, unfortunately, not directly comparable in many cases. The Census Bureau provides a description of the two systems at www.census.gov/epcd/www/naics.html.

[11]Confining the analysis to the 1980 and 1990 data produces estimates similar to what I report here.

area-level employment figures for each industry are computed by aggregating the county-level figures reported in the CBP.[12] Additional data on US metropolitan areas - resident population, density, unemployment - is derived from the USA Counties 1998 on CD-ROM (US Bureau of the Census (1999)) for the years 1980 and 1990 and the US Census Bureau and Bureau of Labor Statistics for 2000.[13]

A few summary statistics characterizing the data appear in the Appendix. In particular, Table A1 reports the college employment fraction for the top and bottom 10 industries (taken across all city-year observations for each industry) which should help to provide some idea about the degree of heterogeneity in the levels of human capital across industries within the manufacturing sector. Within this particular sample, college rates range from less than 0.07 for Logging (Census code 230) and Leather Tanning and Finishing (220) to nearly 0.5 for Drugs (181) and Guided Missiles, Space Vehicles, and Parts (362). Table A2 presents statistics for a some selected individual-level characteristics used in the wage regressions below.

3 Results

3.1 Human Capital and Localization

As noted in the Introduction, indirect evidence suggests that clusters of industry ought to be characterized by larger stocks of human capital. In this section, I take a direct look at

[12] Occasionally, employment is reported in the CBP as a range to adhere to disclosure regulations. Although 12 such ranges are described in the CBP documentation (0-19, 20-99, 100-249, 250-499, 500-999, 1000-2499, 2500-4999, 5000-9999, 10000-24999, 25000-49999, 50000-99999, 100000 or more), the largest two categories did not appear for any of the industries considered here. Where employment is reported in this way, I estimate by taking the midpoint of the range.

[13] County-level population is derived from the 2000 Census at www.census.gov. County-level unemployment data is derived from the Bureau of Labor Statistics' Local Area Unemployment Statistics files at www.bls.gov/lau/home.htm.

this relationship by specifying the average human capital in industry i of city c in year t, H_{ict}, as

$$H_{ict} = \mu_i + \mu_c + \mu_t + \theta \log\left(\text{Emp}_{ict}\right) + \beta \mathbf{Z}_{ct} + \epsilon_{ict} \tag{1}$$

where Emp_{ict} is the industry's total employment in city c at time t; $\mathbf{Z}_{ct}$ is a vector of time varying city-level characteristics (described below) that may influence the extent to which all industries in the same city employ skilled or unskilled workers; μ_i, μ_c, and μ_t are industry-, city-, and time-specific fixed effects included to account for exogenous differences in human capital across sectors, locations, and years; and ϵ_{ict} is a residual.

Again, four measures of city-industry human capital are considered: (i) average years of schooling, (ii) average years of (potential) work experience, (iii) the fraction of college graduates in total employment, and (iv) the fraction of total hours worked accounted for by college graduates. Collectively, these four encompass the most commonly studied measures in the literature on local human capital externalities. Note, the basic intent behind the estimation of (1) is merely to compute partial associations to see how city-industry human capital and scale are related. The equation is *not* intended to quantify the causal association between the two.

Because the dependent variables are calculated based on varying numbers of Census observations, they likely involve some sampling error which, itself, is inversely associated with the number of observations used to compute H_{ict}. A college employment fraction based on 5 observations, for example, likely involves greater variance than one based on 1000. To account for this aspect of the human capital measures, I estimate (1) using a weighted (or generalized) least squares procedure where the weights are given by the number of observations used in the calculation. Hence, a city-industry with 5 observations is given smaller weight than one with 1000.

Two specifications of (1) are considered. In the first, I drop the vector of time varying city-level covariates, $\mathbf{Z}_{ct}$, in an attempt to focus purely on the association between log industry employment and city-industry human capital. The resulting localization estimates, $\hat{\theta}$, are reported in the rows labeled I in Table 1. While it is evident from the goodness-of-fit statistics that the fixed effects collectively account for a large fraction of the total variation in each dependent variable, the coefficient on log industry employment is significantly positive in each of the four cases. To be sure, the implied magnitudes are somewhat modest: a 1 standard deviation increase in an industry's employment[14] within a metropolitan area is accompanied by an increase of 0.04 years in mean schooling, 0.7 years in mean experience, and 1 percentage point in either college fraction. Nevertheless, they indicate that concentrations of industry tend to be characterized by greater quantities of human capital.

Of course, because this first specification does not account for a variety of city-level characteristics that likely influence the degree to which industries employ skilled or unskilled workers, I also estimate a second specification in which the vector $\mathbf{Z}_{ct}$ is added back to the regression. Foremost among the quantities included is the 'overall' college share for the metropolitan area (i.e. the fraction of the population 25 or older with a bachelor's degree). A priori, one would expect that a larger supply of college-educated individuals in the local labor market would increase the extent to which industries hire educated workers. In addition, I add two overall scale effects - the logarithms of resident population and population density - since workers with high levels of human capital may have a tendency to cluster in large metropolitan areas (Glaeser (1999)).[15] I further include four population

[14] In these data, the standard deviation of employment taken across all industries is approximately 2.

[15] Metropolitan area density is calculated as a weighted average of constituent county-level densities, where the weights are given by county population shares. This provides a more representative measure of density (i.e. that faced by an average resident) than city-level average density (total city population divided by total city area). It also may help to mitigate the effects of extremely large but sparsely populated counties, such as many in the western US, on the calculations.

fractions - under 18 years of age, older than 64 years of age, non-white, and foreign born - in an attempt to capture the influence of the local demographics on city-industry human capital.[16] These results are reported in specification II of Table 1.

Looking at the coefficients on the city-specific covariates, the one that appears to be the most important is the overall college fraction which enters significantly in every case. Indeed, each of the education-based measures of industry-specific human capital is positively associated with the share of the adult population with a college degree. This, of course, should not come as much of surprise. After all, if employers hire workers from the local population, a larger stock of highly educated individuals should correspond to a larger fraction of highly educated workers across all local industries. Interestingly, when considering the college employment and hours fractions, the coefficients on the overall city-level college rate are essentially equal to unity, implying that a 1 percentage point rise the population with a bachelor's degree is accompanied by a 1 percentage point increase in a typical industry's college share. The relationship between average experience and the overall college fraction, by contrast, is negative which, again, likely reflects the negative association between age/experience and educational attainment in the data. Among the remainder of the variables, none produces a consistently significant coefficient across all four human capital measures, although each variable does enter significantly in at least one of the four instances.

More importantly, however, none of the estimated coefficients on log industry employment changes substantially after controlling for these additional variables. The coefficient for mean years of education does rise somewhat, 0.019 to 0.024, whereas that for mean years of experience drops from 0.35 to 0.34. Yet, the coefficients for the two college shares remain the same as before. Such a finding is particularly interesting because it suggests that, even after accounting for some basic characteristics of the local population (including a direct

[16] Each of these quantities is calculated from the 5 Percent IPUMS samples.

measure its human capital), larger clusters of city-industry employment are associated with higher average levels of human capital.

These estimates, it should be further noted, change little when the sample is confined to city-industry-year observations for which there are at least 25 Census observations used in the human capital calculations (thereby eliminating the noisiest observations). Doing so reduces the size of the sample from 31460 to 10767. The resulting coefficients (standard errors) on log industry employment from specification II turn out to be very similar to those reported in Table 1: 0.032 (0.013) for mean years of education, 0.34 (0.04) for mean years of experience, 0.008 (0.002) for the college employment share, 0.007 (0.002) for the college hours share.

How uniform are these patterns across industries? Estimates from specification II in which the log industry employment coefficients have been permitted to vary by sector appear in Table 2 for some selected industries. As one might expect, there is a wide range of coefficients observed for each of the four dependent varaibles. For either college share, for instance, the largest association with log employment is that for (non-newspaper) Printing and Publishing (Census code 172) where the coefficient, 0.025, implies a 2.5 percentage point rise in the college fraction as employment (approximately) doubles in the cross section. At the other extreme is Primary Aluminum (code 272), for which the coefficient suggests a decrease of roughly 1 percentage point as employment doubles.

Despite this heterogeneity, the industry-specific estimates suggest precisely the same basic conclusion drawn from the pooled sample. Indeed, given that most of the 77 industries under consideration produce positive coefficients on log city-industry employment – 50 for the college employment share, 51 for the college hours share, 74 for mean experience, and 39 for mean years of education[17] – increases in human capital with localization appears to

[17]In each case, more than half of the positive coefficients differ significantly from zero: 27 for the college employment share, 26 for the college hours share, 62 for mean experience, 30 for mean years of schooling.

be a fairly robust finding.

3.2 Wages, Human Capital, and Localization

Given that human capital scales positively with localization, I turn to my primary question: do localization economies, to any significant degree, reflect human capital externalities? To address this matter, I estimate a series of hedonic wage regressions of the following general form:

$$w_{ict}^{j} = \mu_i + \mu_c + \mu_t + \beta_t \mathbf{X}_{ict}^{j} + \gamma \mathbf{Z}_{ct} + \delta \mathbf{M}_{ict} + \epsilon_{ict}^{j} \tag{2}$$

where w_{ict}^{j} is the log hourly wage of worker j of industry i, city c, in year t; μ_i, μ_c, and μ_t are again industry-, city-, and time-specific fixed effects;[18] $\mathbf{X}_{ict}^{j}$ is a vector of person-specific observable characteristics, including years of education, four educational attainment indicators (no high school, some high school, some college, college), years of education interacted with these four indicators, a quartic in potential experience, eight occupation dummies[19], and a marital status dummy; $\mathbf{Z}_{ct}$ represents a vector of city-time varying characteristics, including log resident population, log population density, the overall college fraction, the unemployment rate, and an estimate of the unionization rate; $\mathbf{M}_{ict}$ is a vector consisting of combinations of city-industry human capital and employment; and ϵ_{ict}^{j} is a residual. Notice, the vector $\mathbf{X}_{ict}^{j}$ is specified with a time-varying set of coefficients to reflect changes in the return to various characteristics (e.g. educational attainment) over time.[20]

[18] These are intended to account for, among other things, the influence of time-invariant city-level amenities (e.g. coastal location, climate) on earnings.

[19] Occupations include Professional-Technical; Managers; Clerical; Sales; Craftsmen; Operatives; Service; and Laborers.

[20] Recall, since the sample is restricted to white males, there is no need to account for parameter heterogeneity based on race or gender.

Because the human capital variables that enter (2) through $\mathbf{M}_{ict}$ must be estimated from Census samples, I eliminate all city-industries involving fewer than 25 observations. Doing so should reduce the error inherent in these regressors and, thus, any potential bias they generate.

I consider several specifications of (2) in which different combinations of a worker's own city-industry scale and human capital are included. The overarching goal is to see how the estimated association between wages and city-industry employment (i.e. the localization 'effect') changes once we condition on human capital. Hence, because the intent of this equation is *not* to estimate a causal relationship between either own-industry employment or city-industry human capital and wage earnings, but instead to see how the estimated associations change once the other regressor is included, I estimate (2) by ordinary least squares.[21] This approach is similar to the one used by, for example, studies of the employer size-wage effect in the labor literature whereby series of wage regressions are estimated to determine the stability of the employer-size coefficient to the inclusion of additional regressors (e.g. Troske (1999)). Results are summarized in Table 3.[22]

Specification I, in which the vector M_{ict} only includes the logarithm of a worker's own-industry employment, demonstrates the standard localization result: after conditioning on a variety of person-specific observable characteristics, there is a significantly positive association between a worker's wage and the extent of his local industry. The implied elasticity from the point estimate (0.028) indicates that a 1 standard deviation increase in log city-industry employment (i.e. approximately 2) tends to be accompanied by a 5.6 percent increase in a worker's hourly earnings. Again, such a value is not inconsistent with what previous research has found (e.g. Henderson (1986), Wheeler (2004)).

[21] Heteroskedasticity-consistent standard errors, adjusted for correlation within metropolitan areas, are computed.

[22] To save space, I have suppressed all of the coefficients from the personal and city-time varying covariates, β_t and γ from Table 3.

The next three specifications add human capital to the regression, but drop a worker's own-industry employment in an effort to gain some sense of how important intra-industry human capital externalities might be. Looking at the columns labeled II, III, and IV, one can see that, across all four variables, the coefficients suggest statistically significant and economically important magnitudes. A 1 year increase in mean education (within a worker's own industry), for example, correlates with a nearly 7 percent increase in wages, whereas a 1 year increase in mean experience is associated with a 0.7 percent rise in wages. These relative magnitudes are similar (at least qualitatively) to Rauch's (1993) findings using city-level education and experience. The estimates also suggest that a 1 percentage point increase in either the college employment fraction or college hours fraction is accompanied by a 0.4 percent increase, approximately, in hourly wages.

Although not reported in the table, the estimated coefficient for the *overall* college rate (among the adult resident population) turns out to be about three times as large, 1.1, as these intra-industry college coefficients. Such a figure, again, is close to what Moretti (2004a) reports across a wide array of statistical specifications.[23] Interestingly, this particular magnitude for the overall college rate does not depend on whether city-industry human capital appears in the regression. The estimated wage gains tied to 'city-level' human capital, therefore, do not appear to be industry-specific in nature. That is, the significance of city-level human capital with respect to wage levels documented in previous work does not stem from the omission of city-industry human capital. Rather, there seem to be strong, positive associations between wages and both human capital measures (i.e. city-level and city-industry-level) that are reasonably independent of each other.

This finding suggests that, although there do seem to be intra-industry human capital externalities, the benefits of greater aggregate human capital are not confined to workers employed in the same industry. Moreover, given the differences in magnitudes, externalities

[23] As mentioned in the Introduction, Moretti's (2004a) estimates range primarily between 1 and 1.5.

arising from city-level human capital appear to be stronger than those emanating from industry-specific human capital.

To what extent do either the localization effects or estimated intra-industry human capital externalities change once both are added to the regressions? The final three columns of Table 3 (labeled V, VI, and VII) report coefficients from specifications in which log own-industry employment, mean years of experience, and one of the three education variables are all included in the model. What they reveal, interestingly, is only a modest change in any of the coefficients. To be sure, there is some drop in all of the estimates just as one would expect given the positive human capital-industry employment relationship documented above. However, the changes are small. The estimated localization elasticity, for example, drops from 0.028 to 0.024 when experience and either college fraction are added. Similarly, the estimated magnitudes of the human capital externalities are altered very little once a worker's own-industry employment is accounted for. The college fraction coefficients, for example, fall from 0.39 to 0.33 once employment is included. These findings indicate that, although positively associated, localization economies and human capital externalities appear to be reasonably distinct economic phenomena.[24] The correlation between localization and wages, therefore, is largely *not* a reflection of human capital spillovers.

3.3 Industry-Specific Estimates

Just as with the relationship between localization and human capital, there are likely to be differences in the estimated localization and human capital 'effects' across industries. To gain some idea about the extent of this inter-industry heterogeneity, I have re-done all of the estimation allowing the coefficients on log own-industry employment and each human capital measure to vary by industry. A sample of the resulting coefficients is provided in

[24] Moretti (2004b) also reports evidence that human capital externalities do not seem to be a manifestation of some agglomeration effect on productivity.

Table 4 for the same set of industries reported in Table 2. To be concise, I have limited the reported output to the coefficients on log own-industry employment, the college employment share, and mean years of experience from a regression in which all three are included as regressors.[25] As noted below, these estimates do not vary substantially from those that arise when these variables are entered into separate regressions.

Localization coefficients, given in the first column of results, do indeed show some variation, ranging from 0.2 for Logging (industry 230) and 0.063 for Cement, Concrete, Gypsum, and Plaster Products at the top end of the distribution to roughly -0.007 for Apparel and Accessories (151) and Miscellaneous Paper and Pulp Products (161) at the bottom. Of the 77 coefficients, 65 are positive (39 significant). Looking at the college fraction and mean years of experience, there is similar variation in the reported coefficients. The college fraction estimates fall between 1.9 (Apparel, 151) and -0.46 (Sawmills, 232) with a total of 63 positive (43 significant). Those for experience range between 0.022 (Apparel, 151) and -0.02 (Logging, 230) with 55 positive (19 significant).

In all, including both human capital and own-industry employment in the same regression (specification VI from Table 3) changes the estimated magnitudes of the coefficients only very little relative to what is produced from separate regressions (specifications I and III from Table 3): mean changes (standard deviation) are -0.0005 (0.014) for log own-industry employment, -0.015 (0.16) for the college employment fraction, -0.002 (0.003) for mean experience. Given the average magnitudes of these coefficients of 0.025, 0.41, and 0.006, these changes tend to be reasonably small in percentage terms, which further underscores the idea that localization economies and human capital externalities represent reasonably distinct mechanisms.

Do industries characterized by large localization 'effects' also have large human capital

[25] Similar inferences can be drawn from the results on mean years of education and the college hours fraction.

externalities? To answer this question, I correlated the estimated coefficients on log own-industry employment, the city-industry college employment share, and city-industry mean years of experience across the 77 industries in the sample. The results show that the estimated localization parameters tend to be negatively associated with the coefficients on both of these two human capital measures: -0.1 (not significantly different from zero) for the college share and -0.4 (significantly non-zero at 1 percent) for mean years of experience. The two estimated human capital externalities, as it happens, are positively and significantly associated, 0.28 (p-value = 0.01).[26] This latter result seems intuitively reasonable in that industries which benefit from the concentration of one type of human capital ought to benefit from another as well.

More importantly, the first two correlations suggest that industries that experience large wage gains from localization do *not* observe large wage gains due to human capital and vice versa. If anything, the results suggest that industries exhibiting large wage gains from localization tend to experience small (or even negative) wage associations with human capital. This insight may help to account for the insensitivity of the own-industry employment and intra-industry human capital coefficients to the presence of the other variable in the wage regressions above. Although city-industry employment and human capital are directly related, the positive association between wages and localization seems to be driven by industries for which human capital externalities are not important. At the same time, human capital externalities are most apparent in industries which do not exhibit strong localization effects. Such inter-industry heterogeneity, then, reinforces the conclusion that localization economies and intra-industry human capital externalities seem to represent distinct phe-

[26] These correlations are based on coefficients from the specification in which industry employment and both human capital measures are included together. The correlations using coefficients estimated from two separate regressions (employment added to one, human capital added to the other) produce similar correlations: -0.15 employment-college share; -0.36 employment-mean experience; 0.23 college share-mean experience.

nomena.

3.4 Education Group-Specific Estimates

One of the important findings reported by Moretti (2004a) is a difference between the estimated magnitudes of human capital externalities across workers of different educational attainment categories. In particular, workers with less education tend to experience more sizable increases in their wages given an increase in the relative supply of college graduates than do more educated workers. Such a finding is consistent with both standard marginal productivity arguments based on supply and demand as well as learning effects as described by Glaeser (1999).[27]

This section considers whether the localization and human capital results documented above show systematic variation across education groups. It may be, for instance, that the positive association between city-industry size and wages is more influenced by the stock of human capital for some groups than others. Indeed, knowledge spillovers may be more important among, say, the college-educated than high school dropouts. In this case, after controlling for city-industry human capital, the estimated localization coefficient should drop substantially for college graduates but very little for high school dropouts.

To investigate this matter, I estimate (2) allowing the human capital and localization coefficients to vary across five educational categories: no high school, some high school, high school, some college, college. Results from the same series of specifications reported in Table 3 are reported by educational category in Table 5. Beginning with the estimated localization parameters, one can see that there is remarkable consistency, at least across the top four groups which produce coefficients close to the 0.028 benchmark estimated from the pooled sample. Workers in the no-high-school category still show a positive association

[27]The less educated, for example, might have the most to gain (in terms of both skill acquisition and wage growth) from interacting with more educated workers.

(albeit insignificant) with log own-industry employment, but the magnitude (0.013) is less than half of what is estimated for the other four groups. Looking at the estimated human capital externalities in the next three specifications (II - IV), a result similar to Moretti's (2004a) is discernible. There is a slight decrease in the size of the human capital coefficients as we move from the no-high-school group to the some-college group.

These findings, however, do show an increase at the top end of the distribution, particularly when considering either college share. This result may, in part, reflect a positive association between either college share and the extent to which industries use skill-biased technologies.[28] Of course, because workers with lower levels of educational attainment appear to gain from higher college shares too, any such technological differences accompanying larger college fractions do not seem to constrain the earnings of the less-educated.

How do these results change once both log own-industry employment and city-industry human capital are both included in the regression? As before, the magnitude of each effect (shown in specifications V - VII) drops slightly just as one would expect given the positive association between employment and human capital. However, with the exception of the two college fractions for the no high school category, the amount by which each coefficient falls is strikingly similar across education groups. Thus, the extent to which localization effects are driven by human capital externalities (and vice versa) seems to be small for workers of all levels of educational attainment.

[28] Acemoglu (2002), for example, argues that the rising supply of college educated workers in the US has spurred producers to adopt skill-complementing technologies (e.g. computers). This idea suggests that there should be a positive association between the extent to which employers hire college-educated workers and the degree to which they utilize skill-biased technologies.

4 Concluding Discussion

This paper has reported evidence that human capital levels, defined in terms of educational attainment and work experience, tend to rise as an industry's scale within a metropolitan area increases. They do so even after accounting for a variety of city-level characteristics, including the general level of education across the resident population. Yet, in spite of this relationship, the boost in wage earnings associated with this rise in human capital does not account for the widely established result of localization economies. The positive association between city-industry scale and hourly earnings remains largely unaltered after accounting for human capital.

What might these results reveal about the nature of the productivity gains tied to the geographic concentration of industry? Human capital externalities, of course, most closely resemble the first of Marshall's (1920) explanations for localization economies mentioned in the Introduction: knowledge spillovers. Although not usually framed in the context of externalities tied directly to the supply of highly educated (or experienced) workers, one could certainly view intra-industry knowledge spillovers as a function of human capital externalities, particularly if increases in the stock of local human capital lead to increases in the amount of knowledge that is generated and exchanged within a local market.[29] From this perspective, evidence of human capital externalities may be interpreted as evidence of Marshallian knowledge spillovers.

Given such an interpretation, the results suggest that localization economies are only in small part a function of knowledge spillovers. While clearly important both economically and statistically – recall, a 10-percentage point rise in the college fraction correlates with a 4 percent increase in hourly wages on average – the level of human capital within a worker's own city-industry does not account for much of the residual boost to labor earnings

[29] Simon and Nardinelli (2002, p. 62) express a similar view, noting that "cities with higher average levels of human capital are also likely to enjoy greater knowledge spillovers."

associated with city-industry scale. Independent of whether one conditions on city-industry education and experience, the estimated magnitude of the wage-localization association is strikingly stable.

The findings also suggest that characteristics that are strongly tied to average human capital, such as the type of physical capital used in production, might not explain localization economies either. Indeed, given industry- and plant-level evidence indicating that producers who utilize more skilled labor also utilize more sophisticated technologies, such as computer aided design and engineering, lasers, and robotics (e.g. Autor et al. (1998), Doms et al. (1997), Troske (1999)), one might hypothesize that localization effects on productivity are driven, in part, by the use of more productive capital.[30] Few studies of the localization phenomenon have attempted to account for the influence of this type of technological difference on productivity. Such differences may, therefore, be an important omitted variable whose influence on wages and productivity are picked up by industrial agglomeration. However, if one assumes that city-industry human capital serves as a reasonable proxy for technological sophistication, these results offer little support for this conjecture.

Of course, such conclusions are to be taken with some caution. As noted above, because intra-industry knowledge spillovers need not be tied to education and experience per se, human capital is likely an imperfect measure of knowledge spillovers at best. Similarly, it provides a less-than-completely desirable (indirect) measure of physical capital sophistication across the producers within a given city-industry. Hence, more direct evidence on both of these mechanisms is needed before their roles in generating localization effects can be assessed. This may be an interesting avenue for future work.

On a more basic level, research examining why clusters of industry tend to draw more educated workers may also prove useful in better understanding the localization phenomenon.

[30] Acemoglu (1996) shows how this relationship can be derived in a simple theoretical framework based on firm-worker matching.

This paper has merely taken first step in attempting to establish this particular empirical result, but has not offered an explanation. One possibility is that large concentrations of industry offer greater learning opportunities for workers. If the workers who gain the most from these learning opportunities happen to be the highly educated, one would expect to see results like those reported here.[31] A similar result could be derived assuming that localization lowers search costs for workers looking for jobs, again to the extent that highly educated individuals have the most to gain from efficient matching. Sorting out these matters may also enhance our understanding of how localization influences economic outcomes.

[31] This explanation mimics Glaeser's (1999) hypothesis regarding the concentration of educated workers in large urban areas.

Table 1: Human Capital and Localization

Pooled Manufacturing

Variable	Mean Years of Education I	Mean Years of Education II	Mean Years of Experience I	Mean Years of Experience II	College Emp. Fraction I	College Emp. Fraction II	College Hours Fraction I	College Hours Fraction II
Log Industry Employment	0.019 (0.01)	0.024 (0.01)	0.36 (0.03)	0.35 (0.03)	0.006 (0.001)	0.006 (0.001)	0.006 (0.002)	0.006 (0.001)
Overall College Fraction	–	6.2 (0.7)	–	-8.9 (3.2)	–	1.15 (0.07)	–	1.19 (0.07)
Log Resident Population	–	-0.03 (0.3)	–	3.7 (1.2)	–	0.04 (0.02)	–	0.03 (0.02)
Log Population Density	–	-0.05 (0.2)	–	-0.66 (0.7)	–	-0.03 (0.02)	–	-0.02 (0.02)
Proportion Under 18	–	-4.5 (1.6)	–	-6.8 (5.4)	–	-0.13 (0.15)	–	-0.12 (0.2)
Proportion Over 64	–	-1.2 (2.8)	–	2.3 (8.1)	–	0.4 (0.17)	–	0.41 (0.2)
Proportion Female	–	2.9 (3.2)	–	57.7 (12.2)	–	-0.36 (0.28)	–	-0.4 (0.3)
Proportion Foreign Born	–	-2.3 (0.8)	–	-3.3 (3.9)	–	0.2 (0.08)	–	0.2 (0.08)
Proportion Non-White	–	-0.98 (0.31)	–	0.49 (1.2)	–	0.05 (0.02)	–	0.05 (0.02)
R^2	0.77	0.78	0.42	0.43	0.72	0.73	0.71	0.73

Note: 31460 observations. Dependent variable is human capital. All regressions include industry-, city-, and time-specific fixed effects and are weighted by the number of city-industry observations used to calculate the dependent variable. Heteroskedasticity-consistent standard errors adjusted for within-city correlation are reported in parentheses.

Table 2: Human Capital and Localization

Selected Industry-Specific Results

		Dependent Variable			
Census Code	Industry	Mean Years of Education	Mean Years of Experience	College Emp. Fraction	College Hours Fraction
110	Grain Mill Products	0.21 (0.04)	0.3 (0.15)	0.013 (0.004)	0.012 (0.004)
111	Bakery Products	-0.09 (0.02)	0.64 (0.09)	-0.0001 (0.003)	0.00006 (0.003)
120	Beverage Industries	0.13 (0.02)	0.72 (0.1)	0.014 (0.003)	0.015 (0.003)
151	Apparel and Accessories	-0.24 (0.01)	0.58 (0.04)	0.0004 (0.001)	0.0006 (0.001)
161	Misc. Paper and Pulp Products	-0.03 (0.02)	0.46 (0.1)	0.003 (0.003)	0.004 (0.003)
172	Printing and Publishing, Except Newspaper	0.17 (0.008)	-0.04 (0.04)	0.025 (0.001)	0.026 (0.001)
181	Drugs	0.09 (0.01)	0.08 (0.06)	0.002 (0.002)	0.002 (0.002)
200	Petroleum Refining	0.08 (0.02)	0.23 (0.1)	-0.001 (0.003)	-0.001 (0.003)
201	Misc. Petroleum and Coal Products	0.09 (0.05)	0.59 (0.24)	0.01 (0.007)	0.011 (0.007)
230	Logging	0.16 (0.04)	0.68 (0.19)	0.003 (0.006)	0.003 (0.006)
232	Sawmills	-0.14 (0.04)	0.23 (0.19)	-0.004 (0.005)	-0.003 (0.006)
242	Furniture and Fixtures	-0.13 (0.01)	0.55 (0.05)	0.002 (0.001)	0.002 (0.001)
251	Cement, Concrete, Gypsum and Plaster	0.1 (0.03)	0.08 (0.13)	-0.004 (0.004)	-0.005 (0.004)
271	Iron and Steel Foundries	-0.01 (0.03)	0.38 (0.11)	0.006 (0.003)	0.007 (0.003)
272	Primary Aluminum	-0.11 (0.03)	0.8 (0.12)	-0.009 (0.004)	-0.01 (0.004)
281	Cutlery and Handtools	-0.03 (0.03)	0.49 (0.12)	0.002 (0.003)	0.002 (0.004)
291	Metal Forgings	-0.004 (0.02)	0.44 (0.09)	-0.002 (0.003)	-0.003 (0.003)
310	Engines and Turbines	-0.01 (0.03)	-0.03 (0.11)	-0.002 (0.003)	-0.004 (0.003)
312	Construction Machines	0.08 (0.02)	0.1 (0.08)	0.006 (0.002)	0.005 (0.002)
322	Electronic Computing Equipment	0.09 (0.01)	0.12 (0.05)	0.009 (0.001)	0.009 (0.001)
340	Household Appliances	-0.05 (0.02)	0.77 (0.1)	-0.003 (0.003)	-0.003 (0.003)
341	Radio, TV, Comm. Equipment	0.11 (0.01)	0.22 (0.05)	0.011 (0.001)	0.011 (0.002)
360	Ship Building and Repair	0.07 (0.02)	0.52 (0.08)	0.005 (0.002)	0.004 (0.002)
361	Railroad Locomotives	0.11 (0.03)	0.12 (0.15)	0.014 (0.004)	0.014 (0.004)
390	Toys and Sporting Goods	-0.04 (0.02)	0.11 (0.1)	0.006 (0.003)	0.007 (0.003)

Note: 31460 observations. Coefficients on log industry employment (by industry) from specification II of Table 1. All regressions are weighted by the number of city-industry observations used to calculate the dependent variable. Heteroskedasticity-consistent standard errors adjusted for within-city correlation are reported in parentheses.

Table 3: Wage Regressions

	Specification						
	I	*II*	*III*	*IV*	*V*	*VI*	*VII*
Log Own-Industry Employment	0.028	–	–	–	0.023	0.024	0.024
	(0.002)				(0.002)	(0.002)	(0.002)
Mean Years of Education	–	0.068	–	–	0.062	–	–
		(0.004)			(0.004)		
Mean Years of Experience	–	0.009	0.007	0.007	0.007	0.005	0.005
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
College Emp. Fraction	–	–	0.39	–	–	0.33	–
			(0.04)			(0.03)	
College Hours Fraction	–	–	–	0.38	–	–	0.32
				(0.03)			(0.03)

Note: 176084 observations. Dependent variable is log hourly wage. Heteroskedasticity-consistent standard errors adjusted for within-city correlation are reported in parentheses.

Table 4: Selected Industry-Specific Estimates

Census Code	Industry	Log Own-Industry Employment	College Emp. Fraction	Mean Years of Experience
110	Grain Mill Products	0.005 (0.02)	0.16 (0.14)	0.003 (0.006)
111	Bakery Products	0.039 (0.01)	0.24 (0.24)	-0.003 (0.006)
120	Beverage Industries	0.042 (0.01)	0.33 (0.1)	0.001 (0.004)
151	Apparel and Accessories	-0.006 (0.006)	1.9 (0.2)	0.022 (0.003)
161	Misc. Paper and Pulp Products	-0.007 (0.01)	0.45 (0.18)	0.008 (0.005)
172	Printing and Publishing, Except Newspaper	0.058 (0.004)	0.0001 (0.06)	-0.001 (0.002)
181	Drugs	0.017 (0.006)	0.44 (0.08)	0.016 (0.006)
200	Petroleum Refining	0.018 (0.009)	0.46 (0.11)	0.015 (0.005)
201	Misc. Petroleum and Coal Products	0.056 (0.06)	-0.04 (0.4)	-0.009 (0.01)
230	Logging	0.21 (0.05)	-0.09 (0.8)	-0.02 (0.01)
232	Sawmills	0.031 (0.04)	-0.46 (0.6)	-0.008 (0.01)
242	Furniture and Fixtures	0.027 (0.005)	0.67 (0.15)	-0.002 (0.003)
251	Cement, Concrete, Gypsum and Plaster	0.063 (0.016)	-0.15 (0.22)	-0.005 (0.005)
271	Iron and Steel Foundries	0.027 (0.01)	0.08 (0.33)	0.001 (0.004)
272	Primary Aluminum	0.01 (0.01)	0.33 (0.16)	0.014 (0.005)
281	Cutlery and Handtools	0.021 (0.014)	0.4 (0.24)	0.003 (0.005)
291	Metal Forgings	0.036 (0.009)	-0.35 (0.23)	0.003 (0.005)
310	Engines and Turbines	-0.003 (0.01)	0.15 (0.16)	-0.003 (0.004)
312	Construction Machines	0.049 (0.007)	-0.01 (0.1)	0.005 (0.003)
322	Electronic Computing Equipment	0.012 (0.004)	0.46 (0.06)	0.007 (0.003)
340	Household Appliances	0.017 (0.013)	0.47 (0.25)	0.003 (0.005)
341	Radio, TV, Comm. Equipment	0.014 (0.005)	0.43 (0.05)	0.008 (0.003)
360	Ship Building and Repair	0.033 (0.008)	0.21 (0.2)	-0.001 (0.004)
361	Railroad Locomotives	0.004 (0.02)	-0.2 (0.4)	0.01 (0.01)
390	Toys and Sporting Goods	0.024 (0.01)	0.63 (0.2)	-0.009 (0.005)

Note: 176084 observations. Depdendent variable is log hourly wage. Heteroskedasticity-consistent standard errors adjusted for within-city correlation are reported in parentheses.

Table 5: Education Group-Specific Estimates

Educ. Group	Spec.	Log Own-Industry Emp.	Mean Years of Educ.	Mean Years of Exp.	College Emp. Fraction	College Hours Fraction
		Independent Variable				
No High School	I	0.013 (0.009)	–	–	–	–
	II	–	0.09 (0.006)	0.015 (0.003)	–	–
	III	–	–	0.01 (0.004)	0.43 (0.09)	–
	IV	–	–	0.01 (0.004)	–	0.4 (0.09)
	V	0.008 (0.007)	0.086 (0.006)	0.014 (0.003)	–	–
	VI	0.008 (0.009)	–	0.009 (0.004)	0.43 (0.1)	–
	VII	0.008 (0.009)	–	0.009 (0.004)	–	0.4 (0.1)
Some High School	I	0.028 (0.003)	–	–	–	–
	II	–	0.08 (0.005)	0.012 (0.002)	–	–
	III	–	–	0.008 (0.002)	0.36 (0.06)	–
	IV	–	–	0.008 (0.002)	–	0.34 (0.06)
	V	0.021 (0.003)	0.075 (0.005)	0.01 (0.002)	–	–
	VI	0.024 (0.003)	–	0.006 (0.002)	0.31 (0.06)	–
	VII	0.024 (0.003)	–	0.006 (0.002)	–	0.29 (0.06)
High School	I	0.027 (0.002)	–	–	–	–
	II	–	0.068 (0.005)	0.009 (0.001)	–	–
	III	–	–	0.006 (0.001)	0.32 (0.04)	–
	IV	–	–	0.006 (0.001)	–	0.3 (0.04)
	V	0.023 (0.002)	0.062 (0.005)	0.006 (0.001)	–	–
	VI	0.024 (0.002)	–	0.004 (0.002)	0.26 (0.03)	–
	VII	0.024 (0.002)	–	0.004 (0.002)	–	0.25 (0.03)
Some College	I	0.032 (0.003)	–	–	–	–
	II	–	0.053 (0.007)	0.007 (0.001)	–	–
	III	–	–	0.006 (0.001)	0.34 (0.05)	–
	IV	–	–	0.006 (0.001)	–	0.33 (0.04)
	V	0.029 (0.003)	0.045 (0.006)	0.005 (0.001)	–	–
	VI	0.029 (0.003)	–	0.004 (0.001)	0.27 (0.04)	–
	VII	0.029 (0.003)	–	0.004 (0.001)	–	0.26 (0.04)
College	I	0.028 (0.004)	–	–	–	–
	II	–	0.064 (0.006)	0.007 (0.002)	–	–
	III	–	–	0.007 (0.002)	0.5 (0.04)	–
	IV	–	–	0.007 (0.002)	–	0.49 (0.04)
	V	0.024 (0.004)	0.057 (0.006)	0.005 (0.002)	–	–
	VI	0.024 (0.004)	–	0.005 (0.002)	0.44 (0.04)	–
	VII	0.024 (0.004)	–	0.006 (0.002)	–	0.44 (0.04)

Note: 176084 observations. Dependent variable is log hourly wage. Heteroskedasticity-consistent standard errors adjusted for within-city correlation are reported in parentheses.

A Appendix

A.1 Census Data

Data used to compute human capital shares for city-industries come from the 5 Percent Samples of the Integrated Public Use Microdata Series (IPUMS). See Ruggles and Sobek et al. (2003). Because the 1990 and 2000 Census do not code educational attainment as years of schooling completed for all individuals, I follow the procedure of Autor, Katz, and Krueger (1998) by imputing years of education from Table 5 of Park (1994). A worker's potential experience is then computed as the maximum of (age-years of education-6) and 0. The calculation of average years of education, average years of experience, the college employment fraction, and the college hours fraction is based on all workers for whom I observe a detailed industry of employment, metropolitan area of residence, positive usual hours worked per week, and positive wage and salary earnings. This corresponds to 753679 observations for 1980, 612957 for 1990, and 641074 for 2000. Average numbers of observations per city-industry follow as 71.5 for 1980 (minimum = 1, maximum = 19377), 55.2 for 1990 (minimum = 1, maximum = 9916), and 52.3 for 2000 (minimum = 1, maximum = 11316). When the sample is restricted to those with at least 25 observations, as in the wage regressions, the averages are 182.3 for 1980, 150.8 for 1990, 141.7 for 2000.

Hourly wages are computed by dividing annual wage and salary earnings by the product of weeks worked and usual hours worked per week. These are converted to real terms using the Personal Consumption Expenditure Chain Type Price Index of the National Income and Product Accounts. To limit the influence of outlier observations on the results, I constrain the sample to workers earning between 1 and 100 dollars per hour (in 2000 dollars) although the results are not sensitive to either cutoff. To facilitate the computations, the wage regressions are based on the 1 Percent IPUMS Samples instead of the 5 Percent Samples. A total of 176084 observations over the three years are used. Average numbers of observations per metropolitan area used in the wage regressions are 394.8 for 1980, 308.3 for 1990, 399.2 for 2000. Averages per city-industry (again, just among observations used in the wage regressions) are 11.3 (minimum = 1, maximum = 2456) for 1980, 11.2 (minimum = 1, maximum = 2921) for 1990, and 15.4 (minimum = 1, maximum = 2185) for 2000.

Metropolitan areas are defined using definitions from 1995 (see US Bureau of the Census (1999)). Because geographic definitions change over time, the metropolitan area codes reported in the IPUMS (particularly those belonging to larger CMSAs) show some spurious variation from year to year. Individuals living in one MSA in 1980, for instance, may be assigned to another within the same CMSA in 1990 or 2000, purely based on definitional changes. For this reason, metropolitan areas that belong to CMSAs are aggregated to the CMSA-level. A total of 204 metropolitan areas are identified for 1980, 196 for 1990, 88 for 2000. These comprise 219 distinct metropolitan areas.

Census industry codes serve as the basis for the industrial classification scheme used in the paper. In most instances, the Census codes correspond to three-digit SIC industries, although some represent two-digit, four-digit, or combinations of three- or four-digit sectors. Because the codes changed between 1980 and 1990, a consistent set of codes were implemented using the crosswalks provided by the U.S. Bureau of the Census. These are described by Barry Hirsch at his website www.trinity.edu/bhirsch. A total of 77 industries appear in the final sample.

A.2 Unionization Rates

Among the regressors included in the wage regressions is an estimate of each metropolitan area's overall rate of union membership. These figures are computed from the state-level unionization rates reported by Hirsch et al. (2001) in the following manner. When a metropolitan area lies completely in a single state, it is assigned the state-level unionization rate. When a metropolitan area spans multiple states (as is often the case), it is assigned a weighted average of its constituent state-level rates, where the weights are given by population shares.

Table A1: College Employment Fractions, 1980-2000

Top and Bottom 10 Industries

Industry	College Employment Fraction
Drugs	0.48
Guided Missiles, Space Vehicles and Parts	0.46
Electronic Computing Equipment	0.43
Industrial and Misc. Chemicals	0.31
Newspaper Publishing and Printing	0.3
Petroleum Refining	0.3
Scientific and Controlling Instruments	0.29
Agricultural Chemicals	0.28
Radio, TV, and Communication Equipment	0.28
Office and Accounting Machines	0.27
Misc. Fabricated Textile Products	0.081
Iron and Steel Foundaries	0.08
Wood Buildings and Mobile Homes	0.077
Screw Machine Products	0.075
Apparel and Accessories, except knit	0.075
Bakery Products	0.072
Knitting Mills	0.072
Meat Products	0.072
Logging	0.067
Leather Tanning and Finishing	0.057

Note: College employment shares for selected industries. Calculations are performed across three years: 1980, 1990, 2000.

Table A2: Summary Statistics

Individual-Level Data

Variable	Mean	Standard Deviation	Minimum	Maximum
Hourly Wage	20.03	11.36	1	99.69
Years of Education	12.71	2.9	0	20
Years of Experience	21.57	12.48	0	59
No High School	0.075	0.26	0	1
Some High School	0.11	0.31	0	1
High School	0.38	0.48	0	1
Some College	0.22	0.41	0	1
College	0.22	0.41	0	1
Resident Population	5018758	5356667	100376	19397717
Population Density	2684.5	3925.9	30.6	16258.1
Overall City College Fraction	0.23	0.06	0.09	0.41
Own-Industry Employment	22185.6	35986.5	9	196751
Own-Industry College Employment Fraction	0.18	0.12	0	0.82
Own-Industry College Hours Fraction	0.19	0.13	0	0.83
Own-Industry Mean Years of Education	12.45	1.1	5.66	16.3
Own-Industry Mean Years of Experience	21.1	2.7	8.3	33.7

Note: 176084 observations. Statistics for all variables (including the three city-level variables and five city-industry variables) are computed as unweighted averages across all individuals.

References

Acemoglu, D. (1996) "A Microfoundation for Social Increasing Returns in Human Capital Accumulation." *Quarterly Journal of Economics.* 111 (3), 779-804.

Acemoglu, D. (2002) "Technical Change, Inequality, and the Labor Market." *Journal of Economic Literature.* 40 (1), 7-72.

Acemoglu, D. and J. Angrist. (2000) "How Large Are Human Capital Externalities? Evidence From Compulsory Schooling Laws." in B. Bernanke and K. Rogoff eds., *NBER Macroeconomics Annual*, Vol. 15, Cambridge, MA: MIT Press, 9-59.

Autor, D., L. Katz, and A. Krueger. (1998) "Computing Inequality: Have Computers Changed the Labor Market?" *Quarterly Journal of Economics.* 113, 1169-1213.

Barron, J., J. Bishop, and W. Dunkelberg. (1985) "Employer Search: The Interviewing and Hiring of New Employees." *Review of Economics and Statistics.* 67(1), 43-52.

Carlino, G. (1979) "Increasing Returns to Scale in Metropolitan Manufacturing." *Journal of Regional Science.* 19 (3), 363-373.

Ciccone, A. and G. Peri. (2004) "Identifying Human Capital Externalities: Theory With Applications." Manuscript, Universitat Pompeu Fabra.

Doms, M., T. Dunne, and K. Troske. (1997) "Workers, Wages, and Technology." *Quarterly Journal of Economics.* 112 (1), 253-290.

Glaeser, E. (1999) "Learning in Cities." *Journal of Urban Economics.* 46, 254-277.

Henderson, V. (1986) "Efficiency of Resource Usage and City Size." *Journal of Urban Economics.* 19, 47-70.

Henderson, V. (2003) "Marshall's Scale Economies." *Journal of Urban Economics.* 53, 1-28.

Hirsch, B., D. Macpherson, and W. Vroman. (2001) "Estimates of Union Density by State." *Monthly Labor Review.* 124 (7), 51-55.

Holmes T. and J. Stevens. (2002) "Geographic Concentration and Establishment Scale." *Review of Economics and Statistics.* 84 (4), 682-690.

Idson, T. (1989) "Establishment Size Differentials in Internal Mobility." *Review of Economics and Statistics.* 71 (4), 721-724.

Jaffe, A., M. Trajtenberg, and R. Henderson. (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics.* 108, 577-598.

Kim, S. (1995) "Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in US Regional Manufacturing Structure, 1860-1987." *Quarterly Journal of Economics.* 110 (4), 881-908.

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

Marshall, A. (1920) *Principles of Economics*. London: Macmillan.

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

Moretti, E. (2004b) "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review*. 94 (3), 656-690.

Oi, W. and T. Idson. (1999) "Firm Size and Wages." in *Handbook of Labor Economics*, vol. 3B, O. Ashenfelter and D. Card eds. New York: Elsevier.

Park, J. (1994) "Estimation of Sheepskin Effects and Returns to Schooling Using the Old and New CPS Measures of Educational Attainment." Princeton Industrial Relations Section Working Paper No. 338.

Rauch, J. (1993) "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities." *Journal of Urban Economics*. 34, 380-400.

Romer, P. (1986) "Increasing Returns and Long Run Growth." *Journal of Political Economy*. 94 (5), 1002-1037.

Ruggles, S. and M. Sobek et al. (2003) *Integrated Public Use Microdata Series: Version 3.0.* Minneapolis: Historical Census Projects, University of Minnesota. http://www.ipums.org.

Simon C. and C. Nardinelli. (2002) "Human Capital and the Rise of American Cities, 1900-1990." *Regional Science and Urban Economics*. 32, 59-96.

Troske, K. (1999) "Evidence on the Employer Size-Wage Premium From Worker-Establishment Matched Data." *Review of Economics and Statistics*. 81 (1), 15-26.

US Bureau of the Census. (1999) *USA Counties 1998 on CD-ROM*. [machine readable data file]. Washington, D.C.: The Bureau.

Wheeler, C. (2004) "Productivity and the Geographic Concentration of Industry: The Role of Plant Scale." Federal Reserve Bank of St. Louis Working Paper No. 024A.