Why Does Educational Attainment Differ Across U.S. States?

Lutz Hendricks

Iowa State University, CESifo, and Center for Financial Studies, Frankfurt*
October 12, 2004

Abstract

The fraction of persons holding a college degree differs nearly two-fold across U.S. states. This paper documents data related to state educational attainment differences and explores possible explanations. It shows that highly educated states employ skill-biased technologies, specialize in skill-intensive industries, but do not pay lower skill premia than do less educated states. Moreover, measures of urbanization and population density are positively related to educational attainment. Theories based on agglomeration economies offer natural explanations for these observations.

Key words: Education, agglomeration. JEL: J24, R11.

1 Introduction

Educational attainment differs substantially across U.S. states. In the five most educated states, 34% of residents hold a college degree compared with only 18% in the five least educated states.[1] The purpose of this paper is to document differences in educational attainment across U.S. states and to investigate possible explanations.

Understanding why educational attainment differs is of obvious public policy interest. States undertake substantial efforts aimed at improving the educational attainment of their residents. Developing a positive theory of local attainment differences is a necessary precursor for analyzing such policies. Differences in educational attainment may also be an important source of variations in income and labor productivity across states.

The paper documents the following main facts:

1. More educated states employ more skilled workers in a given industry.

2. More educated states specialize in skill-intensive industries.

3. Differences in educational attainment across states or metropolitan areas (MAs) are very weakly correlated with skill premia, even within industries.

---

[1]These figures are for persons aged 26 to 60 in 1990. Section 2 describes the data in detail.

These findings suggest that more educated places employ more skill-biased technologies within each industry. I develop a simple model where locations are endowed with industry-neutral differences in technological skill biases. Workers are perfectly mobile so that wage rates are equalized across locations. I show that this simple model accounts for more than 90% of the variation in college attainment observed across state-industry pairs. By contrast, differences in regional skill prices play only a small role in accounting for educational attainment gaps.

The key challenge for understanding why educational attainment differs across U.S. states is therefore to understand the sources of skill biased technology differences. Why do some places employ skill-biased technologies across all industries, given that relative skill prices do not differ from other locations? I argue that agglomeration economies offer a natural explanation. Consistent with this idea, the empirical analysis shows that educational attainment is correlated with a number of measures of agglomeration economies, such as urbanization and population density. A number of agglomeration theories suggest that larger places are more productive because they offer better opportunities for specialized knowledge. Helsley and Strange (1990) argue that larger places permit better matches of worker skills and job characteristics. Abdel-Rahman and Fujita (1990) suggest that larger places sustain more varieties of non-tradeable intermediate inputs, such as business services.[2] I develop models based on these ideas that account qualitatively for the relationship between population size, educational attainment, and skill premia observed in U.S. data.

Even though the focus of this paper is attainment differences across states, agglomeration theories suggest that the appropriate unit of observation is the city, not the state. I therefore present data for metropolitan areas (MAs) as well as states where such data are available.

To my knowledge, this is the first attempt to account for differences in educational attainment across U.S. states. Still, this paper is related to a large body of earlier work. Goldin (1998, 1999) studies the emergence of secondary education in the U.S. around 1920. Her analysis has obvious implications for regional attainment differences, but the focus is on a somewhat different question and on an earlier time period. Bound et al. (2004) study "how the production of college graduates at the state level affects the stock of college-educated workers in the state" (p. 147). Kaboski (2003) studies differences in the returns to education across countries. He develops a model that also accounts for differences education levels. Peri (1998) finds that schooling and Mincerian returns to schooling are positively correlated across U.S. cities and interprets this as evidence of directed technical change.

The idea that cities offer high returns to human capital is prominent in the agglomeration literature. Rauch (1993) presents evidence that wages are higher in more educated cities. Glaeser (1999) and Glaeser and Mare (2001) suggest that cities are more educated because they offer better opportunities for learning. Eaton and Eckstein (1997) develop a model where cities differ in human capital due to learning spillovers that vary across cities. Black and Henderson (1999) offer a model where cities differ in educational attainment because they specialize in different industries. Aside from the focus on states rather than cities, the main innovation in this paper is to show that more educated places feature more skill biased (reduced form) technologies within industries, thus employing more skilled labor without paying lower skill premia.

---

[2]See Duranton and Puga (2003) for a survey of the literature on agglomeration economies.

The paper is organized as follows. Section 2 describes the data and documents differences in educational attainment across U.S. states. Section 2.2 develops a measurement framework that helps interpret the data. The relationship between skill premia and educational attainment is documented in section 3. Section 4 measures the importance of industry specialization for attainment differences. Theories that account for the patterns observed in the data are developed in section 5. The final section concludes.

2   Differences in Educational Attainment Across U.S. States

This section documents how educational attainment differs across U.S. states. The objective is to describe a set of observations that a theory of educational attainment should account for.

Data Sources: The data are drawn from a variety of sources which are described in detail in the appendix. The main data sources are the Decennial Censuses from 1940 to 1990. Earlier years do not contain data on educational attainment.[3] Individuals are dropped if they live in group quarters, if they are younger than age 26 or older than age 60. Age 26 is chosen such that most persons have completed education. Age 60 is chosen so that most persons participate in the labor market. Educational attainment is measured by years of completed schooling. A person with at least 16 years of schooling is classified as holding a college degree. I refer to such persons as "skilled."

The geographic units considered are states and metropolitan areas (MAs). Because of its special status, the District of Columbia is dropped from the sample. The geographic composition of MAs differs between census years. Occasionally, the data are augmented with additional sources that cover finer geographic detail (below the MA level) or that contain information about regional aggregates that cannot be computed from the PUMS data, such as land area. More detail is provided in the Appendix.

From the perspective of some theories that might account for regional attainment differences, cities are the natural geographic units. However, since not all data are available and since sample sizes can be small at the MA level, the main focus of the paper is to understand attainment differences across states. Wherever possible, I present data at the MA level as well as the state level.

Most of the findings reported in this paper are based on 1990 data. Results for 1980 are similar, except where explicitly noted. Results for earlier years are qualitatively similar, but can be quite different quantitatively. To some extent these differences may be due to the much smaller sample sizes or to differences in geographic coverage in earlier years. In early years, when many persons did not graduate from high school, it may also be less appropriate to divide the population in skilled and unskilled workers based on whether persons hold college degrees.

2.1   Basic Observations

The focus of this paper are differences across U.S. states in the fraction of persons holding a college degree. I refer to this fraction as *college attainment*, λ. Table 1 reports that college attainment in the five most educated states averages 34% in 1990. By contrast, in the five

---

[3]Baier et al. (2004) construct state education data at decadal frequency between 1840 and 2000.

least educated states attainment averages only 18%. This is the central fact to be explained: college attainment differs nearly two-fold across U.S. states.

State	Attainment	State	Attainment
MA	35.5	MS	19.1
CT	35.1	NV	18.2
NJ	33.5	KY	17.8
MD	32.1	AR	17.7
CO	31.7	WV	16.4
Mean	33.6	Mean	17.8

Table 1. College attainment across U.S. states.

Persistence: College attainment is quite persistent over time. The correlation coefficient over the period 1940 to 1990 is 0.47. There is no strong evidence that state attainments converge over time. The absolute attainment gap between the most educated states and the least educated states rose slightly between 1940 and 1990. However, since mean attainment increased substantially, the coefficient of variation declined from 0.25 to 0.19. Using mean years of schooling as the measure of education shows much stronger convergence. The gap between the top five states and the bottom five states fell from 3 years in 1940 to 1 year in 1990.

In what follows I examine which variables are correlated with state attainment. The choice of variables is motivated by theories that might account for regional attainment differences.

2.1.1 Demographics

Table 1 aggregates attainment over persons with very different demographic characteristics. This raises the question whether differences in the age/race/sex composition of the population are important for attainment differences across states. To shed light on this question, I decompose attainment differences into the contributions of (i) attainment variations among persons of identical characteristics and (ii) variations in the demographic composition of the population.

Specifically, I divide the population in each state into 5-year age cells, 3 race cells (white, black, other) and 2 sex cells (male and female). For each cell k, I calculate college attainment in state m, denoted by $\lambda_{k,m}$, and the fraction of the state's population in the cell, $\omega_{k,m}$. College attainment in state m may then be written as

$$\lambda_m = \sum_k \omega_{k,m} \; \lambda_{k,m}$$

Denote by $\omega_{k.US}$ the fraction of the U.S. population in cell k and by $\lambda_{k,US}$ the U.S. fraction of college graduates in each cell. To measure the importance of demographic composition, I calculate state m's attainment imposing U.S. attainment within each cell:

$$\hat{\lambda}_m = \sum_k \omega_{k,m} \; \lambda_{k,US}$$

4

By eliminating attainment differences among households with identical characteristics, $\hat{\lambda}_m$ isolates the contribution of demographic weights. Figure 1 plots $\hat{\lambda}_m$ against λ_m for 1990 data. Only a small fraction of state attainment variation is accounted for by demographic weights. Regressing $\hat{\lambda}_m$ against λ_m yields a slope coefficient of only 0.037 ($R^2 = 0.024$).
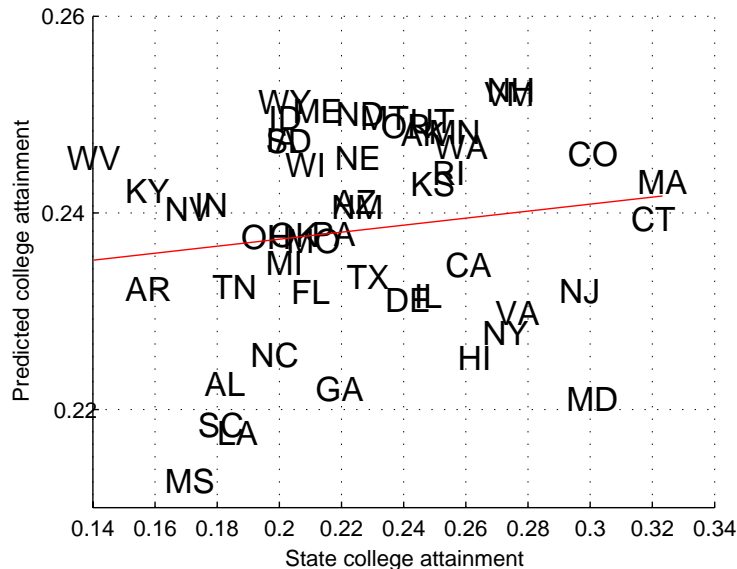

Figure 1: College attainment differences due to demographic structure.

In a similar way the contribution of attainment differences among persons with similar characteristics can be measured. Let $\tilde{\lambda}_m$ denote state m's attainment imposing U.S. cell weights:

$$\tilde{\lambda}_m = \sum_k \omega_{k,US}\ \lambda_{k,m}$$

By eliminating variation in demographic composition, $\tilde{\lambda}_m$ isolates the role of cross-state differences in cell means. Regressing $\tilde{\lambda}_m$ against λ_m yields a slope of 0.93 which is not significantly different from 1 ($R^2 = 0.93$). I conclude that demographic composition accounts for only a small part of the variation in educational attainment across U.S. states.

2.1.2 Supply of Graduates

Bound et al. (2004) study whether states that produce more college graduates also employ more college graduates. They find "at best only a modest link" (p. 147). For completeness, I confirm Bound et al.'s result in my data. I regress state attainment on the number of college graduates per person, which is constructed from the Digest of Education Statistics, 1994 (table 235) for the 1990-91 academic year. Consistent with Bound et al.'s (2004) finding, the relationship is positive, but the R^2 is modest (0.14). Adding a measure of the college skill premium, described more fully in section 2.2, increases the R^2 measure to only 0.15. The coefficient on the skill premium is negative, but not statistically significant (-0.061, $s.e.$ 0.067). One interpretation, which is consistent with other evidence presented below, is that worker mobility across states results in approximate relative factor price equalization.

2.1.3 Agglomeration Effects

In this section, I present evidence that larger or more densely populated places are more educated, suggesting that agglomeration economies may account for regional education gaps.

At the state level, attainment is positively related to measures of urbanization. Regressing state attainment against the fraction of the population living in MAs yields a coefficient of 0.099 (s.e. 0.031, $R^2 = 0.18$). One reason for the positive relationship is that MA residents are substantially more educated than non-MA residents. Figure 2 shows mean state attainment inside and outside MAs for 1990. Two observations are important: (a) Attainment outside of MAs is, on average, 9 percentage points lower than within MAs (18% vs. 27%). (b) States with high MA attainment also have high non-MA attainment. The correlation coefficients between state college attainment and the fraction of the population living in urban areas or in MAs ranges from 0.38 to 0.64 across all years.[4]
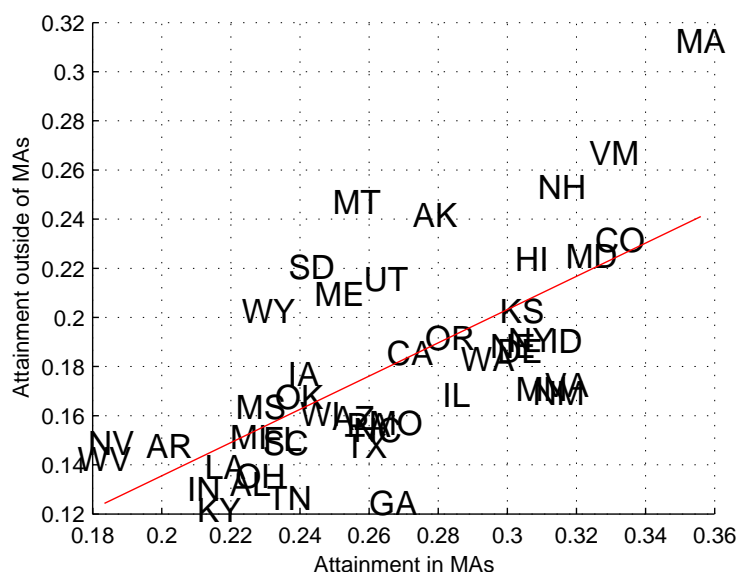

Figure 2: College attainment: MAs versus non-MAs

Among city residents, those living in larger cities are more educated. A regression of MA college attainment on log population yields a coefficient of 0.025 (s.e. 0.006, $R^2 = 0.08$). Population size accounts for 5.1% of the 17.1% attainment gap between the most educated and the least educated cities.[5] The positive relationship between city population and educational attainment has been documented before (e.g., Black and Henderson 1999).

A similar picture emerges for urban versus rural areas. The exact definition of urban areas in the PUMS data varies over time. Typically, places with more than 2,500 residents are classified as urban. Around two-thirds of households live in urban areas in all years with

---

[4]This excludes the years 1960 and 1970 where it is not known whether many households resided in urban or rural areas.

[5]That is, the predicted attainment gap between cities in the 90th versus the 10th percentile of the population size distribution is 5.1%. The observed gap between the 90th versus the 10th percentile of the attainment distribution is 17.1%.

data. Figure 3 shows data for 1990. Again, two observations are important: (a) Attainment in urban areas is 10 percentage points higher than in rural areas (28% versus 18% in 1990). (b) States with high urban attainment also have high rural attainment.

Figure 3: College attainment: rural versus urban areas

To measure the potential role of agglomeration economies, I regress state attainment on various measures of agglomeration in 1990: the fraction of the population living in MAs, the fraction living in urban areas, the average population size of MAs in the state, and the population density in the state. Since the regressors are strongly correlated, the individual regression coefficients are not informative. However, taken together, these regressors account for 33% of the variation in attainment across the 47 states for which data are available. Including Ciccone and Hall's (1996) measure of density increases R^2 to 0.52. These finding suggests that agglomeration economies could account for a substantial share of cross-state variation in educational attainment.

Summary: The data presented in this section suggest that a theory of educational attainment should account for the following facts:

1. College attainment differs by a factor of two across U.S. states.

2. These differences are persistent over time.

3. Attainment is only weakly correlated with the supply of graduates and with the skill premium.

4. Attainment is positively related to measures of agglomeration, such as population size or density.

2.2 Measurement Framework

In this section, I set up a model that helps to interpret the data and to clarify why attainment differs across locations. There are M locations, indexed by m, which could represent states or cities. There are I industries, indexed by i, and a large number of individuals, indexed by j. Good i is produced in location m from skilled labor, $H_{m,i}$, and unskilled labor, $L_{m,i}$ with the constant elasticity of substitution (CES) production function

$$F\left(H_{m,i}, L_{m,i}\right) = A_{m,i} \left[\mu_i \varphi_m \, H_{m,i}^{\rho} + L_{m,i}^{\rho}\right]^{1/\rho} \tag{1}$$

Industries differ in their skill weights μ_i where the weights on unskilled labor are normalized to unity. In addition, the production function allows for industry-neutral differences in skilled labor augmenting technology, φ_m. The motivation for this assumption will become evident when industry-level data are presented in section 4. For most of the analysis, there is no need to impose restrictions on the productivity coefficients $A_{m,i}$.

Profit maximization of price-taking firms implies that factors are paid their marginal products:

$$w_{H,m,i} = p_{m,i} \, \mu_i \, \varphi_m \, A_{m,i} \left[\mu_i \varphi_m \, H_{m,i}^{\rho} + L_{m,i}^{\rho}\right]^{1/\rho - 1} \, H_{m,i}^{\rho-1}$$

$$w_{L,m,i} = p_{m,i} \, A_{m,i} \left[\mu_i \varphi_m \, H_{m,i}^{\rho} + L_{m,i}^{\rho}\right]^{1/\rho - 1} \, L_{m,i}^{\rho-1}$$

where $p_{m,i}$ is the price of good i in place m. Skill premia are given by

$$\frac{w_{H,m,i}}{w_{L,m,i}} = \mu_i \, \varphi_m \, \left(H_{m,i}/L_{m,i}\right)^{\rho-1} \tag{2}$$

In permitting skill premia to differ across industries, the model allows for restrictions in worker mobility between industries. None of the findings depend on this assumption because deviations from relative factor price equalization turn out to be unimportant for understanding attainment gaps across locations.

In the absence of technology differences, the model has familiar Heckscher-Ohlin implications. If relative factor prices are equalized across locations, then a given industry employs the same fraction of skilled labor in every location. If relative factor prices are not equalized, then skill premia are low in places where $H_{m,i}/L_{m,i}$ is high for any given industry.

Empirical estimates place the elasticity of substitution between skilled and unskilled labor between 1.2 and 2 (see Ciccone and Peri 2004 and the references cited therein). Based on these estimates I set $\rho = 0.4$ which implies a substitution elasticity of 1.67.

2.3 Measurement

In measuring labor inputs, it is necessary to account for the possibility that workers in the same education class differ in their labor efficiency. I estimate person j's efficiency using a standard Mincerian earnings equation of the form

$$\ln g\left(e_j, x_j, z_j\right) = \beta_0 + \beta_1 e_j + \beta_2 x_j + \beta_3 x_j^2 + \beta_4 z_j + \varepsilon_j \tag{3}$$

where e denotes years of schooling, x is experience, and z denotes other demographic characteristics (race and sex). Labor efficiency (g) is proxied for using nominal hourly wage rates.

The equation is estimated for the entire U.S. working population, but separately for skilled and unskilled workers. Skilled labor input in industry i is then given by

$$H_{m,i} = \sum_j g\left(e_j, x_j, z_j\right) \ l_j \tag{4}$$

where l_j denotes hours worked and the sum covers all persons working in industry i and location m. A similar equation defines unskilled labor input, $L_{m,i}$. I abstract from the possibility that workers differ in their unmeasured skills. For most of my results this only matters if skilled workers differ from unskilled workers within a given location or industry. In that case, estimated college attainment

$$\eta_{m,i} = H_{m,i} / \left(H_{m,i} + L_{m,i}\right) \tag{5}$$

differs from the true one. For reasons discussed in appendix B, this possibility is not pursued in the analysis.

Conceptually, measuring labor input in efficiency units is more satisfactory than simply counting the number of persons employed. However, quantitatively both measures yield very similar results. The correlation coefficient between η_m and λ_m at the state level is 0.97 and the slope of a regression of η_m on λ_m is 0.92.

It is convenient to work with wage bills, which are directly observed, rather than with wage rates per efficiency unit of labor $(w_{H,m,i}, w_{L,m,i})$ which are not directly observed. Wage bills are defined as

$$W_{H,m,i} = w_{H,m,i} \ H_{m,i} \tag{6}$$
$$W_{L,m,i} = w_{L,m,i} \ L_{m,i} \tag{7}$$

The wage bill ratio is given by

$$\frac{W_{H,m,i}}{W_{L,m,i}} = \mu_i \ \varphi_m \ \left(H_{m,i}/L_{m,i}\right)^\rho \tag{8}$$

This equation is useful for estimating the parameters of the production function. The skill premium may be then be estimated from

$$\frac{w_{H,m,i}}{w_{L,m,i}} = \frac{W_{H,m,i}}{W_{L,m,i}} \frac{L_{m,i}}{H_{m,i}} \tag{9}$$

3   Skill Premia

One way of gauging whether differences in attainment are due to variations in the demand for education or to variations in the supply of educated labor is to study the relationship between skill premia and attainment. The main finding is that skill premia do not systematically vary with college attainment across states or cities. In this section I abstract from multiple industries, but I show below that a similar finding holds within industries.

At the state level, the correlation between skill premia and college attainment is negative in some years and positive in others, but always weak. Figure 4 shows state data for 1990. The correlation coefficient is -0.13 ($+0.04$ in 1980). A 20 percentage point increase in college
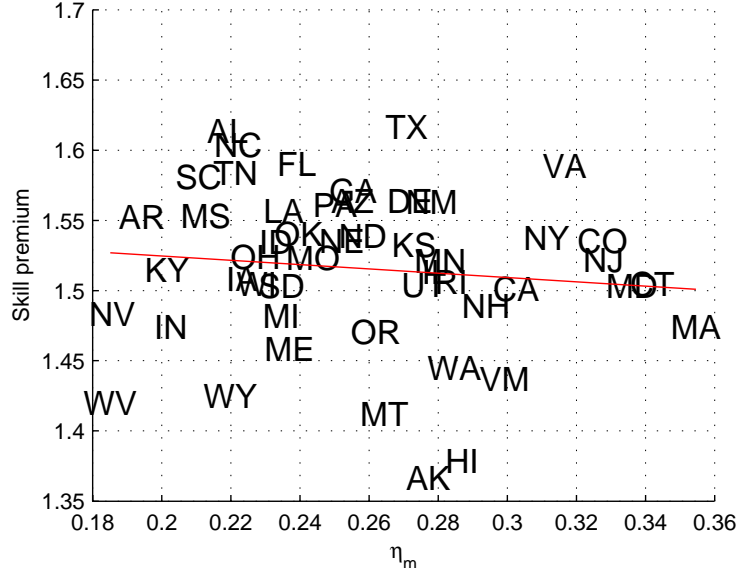
Figure 4: State college attainment and skill premia

attainment is associated with a drop in the skill premium from 1.55 to 1.50. Results at the MA level are similar.

The weak correlation between attainment and skill premia is important. It suggests either that skilled and unskilled labor are very highly substitutable or that highly educated places employ technologies that are relatively skill intensive. The former interpretation is difficult to reconcile with empirical estimates of substitution elasticities. If technologies (φ_m) do not differ across states, then the 1990 data imply a substitution elasticity between skilled and unskilled labor near 30. But empirical estimates place the substitution elasticity around 1.5 (Ciccone and Peri 2004). Even with a substitution elasticity of 4, an increase in the college fraction from 0.2 to 0.4 should be associated with a drop in the skill premium from 1.55 to 1.22, compared with a drop to 1.50 in the data.

An alternative interpretation is that more educated places employ more skill biased technologies. Theories of directed technical change could rationalize this fact (Acemoglu 2002, 2003). Using the measurement framework developed in section 2.2 with a single industry, the value of φ_m for each state or MA may be calculated from

$$\frac{w_{H,m}}{w_{L,m}} = \mu \; \varphi_m \; (H_m/L_m)^{\rho-1} \tag{10}$$

where μ may be normalized to 1.

Figure 5 plots skill weights and college attainments for states in 1990. The correlation coefficient is 0.95. Similar results are obtained for other years or for data at the MA level. Moreover, MAs with growing attainment also experience faster growth of φ_m. For the period 1980 to 1990 the correlation between attainment changes and φ_m changes is 0.74. These observation suggests that differences in the *demand for skilled labor* may be important for understanding educational attainment differences across cities and states.

10

Figure 5: State college attainments and skill weights.

Next, I measure the fractions of college attainment differences that are accounted for by differences in skill premia and skill weights. Inverting (10) yields an expression for college attainment. To quantify the importance of skill premia, I calculate predicted attainment based on the observed skill premia, π_m, but imposing the U.S. average φ:

$$\widetilde{H_m/L_m} = (\varphi_{US} \ / \ \pi_m)^{1/(1-\rho)}.$$

The predicted attainments are essentially uncorrelated with observed attainments, as shown in figure 6 for 1990 MA data. Similar results are obtained for 1980 and for state data. I conclude that differences in skill premia account for a negligible fraction of state or MA attainment gaps.

To quantify the importance of differences in skill weights, I calculate the predicted values of H_m/L_m with the values of φ_m estimated for each MA, but imposing the U.S. average skill premium (π_{US}):

$$\widehat{H_m/L_m} = (\varphi_m \ / \ \pi_{US})^{1/(1-\rho)}.$$

The predicted college attainments are nearly identical with the data values (the correlation is 0.95). Similar results are obtained from state data.

One interpretation is that a model which correctly predicts the values of φ_m can account for more the 90% of the variation in attainment across MAs. This is true, even if the model abstracts from differences in factor prices across MAs. In what follows, the analysis will therefore abstract from obstacles to labor mobility and focus on models that imply relative factor price equalization.[6]

The key problem is then to understand why φ_m differs across locations. Two classes of theories come to mind. One class of theories holds that educated places specialize in

---

[6]This is consistent with Bound et al.'s (2004, p. 147) finding of "at best only a modest link" between the production and employment of college graduates across U.S. states.
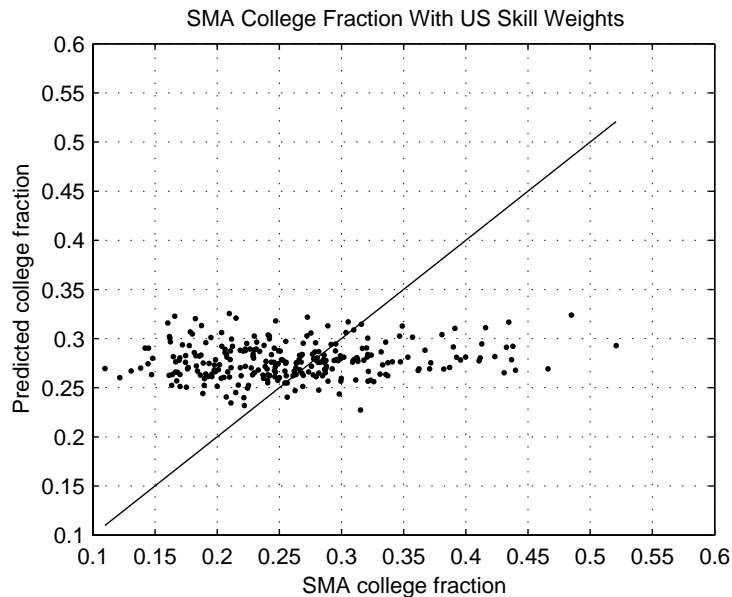
11

Figure 6: Predicted MA attainement with common skill weights

skill intensive industries while employing the same fraction of skilled workers within each industry. A Heckscher-Ohlin model with factor price equalization makes this prediction. A second class of theories holds that locations differ in the skill intensity of technologies within each industry. Therefore, educated places employ more skilled workers in all industries. The next section attempts to measure the relative importance of industry specialization and skill bias differences for understanding college attainment gaps across states.[7]

4 Industry Specialization

With multiple industries, differences in skill weights across locations could reflect differences in industry employment shares. Heckscher-Ohlin trade theory offers an interpretation along these lines. In fact, all of the data reported so far are consistent with a Heckscher-Ohlin model with factor price equalization. Within each industry this model predicts that H/L should be the same in all states, but more educated places should specialize in skill-intensive industries. I investigate these predictions using data on industry employment and wages at the state level for 1990 (results for 1980 are similar). The industry classification is taken from PUMS data and distinguishes roughly 400 industries.

Skilled states employ more skilled labor in all industries: First, I investigate the prediction that $H_{m,i}/L_{m,i}$ is the same for all states for a given industry. Direct inspection of examples shows that this is not the case. Figure 7 plots college attainment in the six largest industries $(\eta_{m,i})$ against state attainment (η_m) for all states with data. In five out of the six industries a strong positive correlation is visible with slopes that are close to one.

To see whether this pattern holds across all industries, I estimate the following equation

---

[7]A third possibility is that workers differ in unmeasured skills across locations. Appendix B discusses why this possibility is not pursued further in the paper.

for each state-industry pair (m, i) via OLS:

$$\eta_{m,i} = \beta_0 + \beta_\eta \eta_m + \gamma D_i + \varepsilon_{m,i}$$

where D_i is a vector of industry dummies. The Heckscher-Ohlin model with factor price equalization predicts that $\beta_\eta = 0$. That is, more educated states should not employ more college graduates in a given industry than do less educated states. The point estimate for β_η in 1990 using the full industry detail available in the Census data is 0.745 ($s.e.$ 0.038). States with 10% more college graduates employ, on average, 7% more college graduates in each industry. This finding is strongly at variance with Heckscher-Ohlin theory with factor price equalization.

Skill premia are not correlated with attainment: One interpretation is that factor price equalization fails to hold in the data. A Heckscher-Ohlin model then predicts, by virtue of (8) and (9), that skill premia should be low in states with high attainment for a given industry. However, in the data, skill premia within industries are essentially uncorrelated with college attainment. To establish this finding, I estimate the skill premium condition (2) under the assumption that all states share the same φ. Specifically, I estimate via OLS

$$\ln \pi_{m,i} = \ln \left(\mu_i / \left(1 - \mu_i \right) \right) + (\rho - 1) \ln \left(H_{m,i} / L_{m,i} \right)$$

where $\ln \left(\mu_i / \left(1 - \mu_i \right) \right)$ becomes the coefficient on an industry dummy. Based on the empirical estimates summarized in Ciccone and Peri (2004), I set $\rho = 0.4$.

Figure 8 plots the industry-state skill premia net of industry dummies, $\pi_{m,i} - \ln \left(\mu_i / \left(1 - \mu_i \right) \right)$, against $\ln \left(H_{m,i} / L_{m,i} \right)$. If all states employed the same technologies within a given industry, the data points should lie along a downward sloping line of slope $\rho - 1$. In the data, by contrast, skill premia are nearly uncorrelated with industry attainment.[8] One interpretation is that more educated states employ more skill biased technologies with higher φ_m. The values of φ_m can be estimated jointly with the values of μ_i from (2) via nonlinear least squares. The resulting correlation coefficient between φ_m and η_m is 0.8 for states in 1990.

Educated states specialize in skill-intensive industries: The results presented so far are inconsistent with a Heckscher-Ohlin model, unless states differ in the values of φ_m. However, this does not imply that industry specialization is not important for state education gaps. Consistent with the model's prediction, educated states specialize in skill-intensive industries. To establish this fact, I rank industries by their estimated values of μ_i and states by their college attainments, η_m. I find that the 10 most educated state employ 46% of labor in top quartile of industries, compared with 39% for the 10 least educated states. In the bottom quartile of industries the 10 most educated states employ 26% of their labor compared with 36% for the 10 least educated states. Results are very similar if industries are ranked by U.S. average education instead of μ_i.

How important is specialization? This raises the question how important industry specialization is for understanding college attainment gaps. I draw on the estimated values of φ_m and μ_i to decompose attainment gaps into the contributions of industry specialization, skill biases (φ_m), and skill premia.

---

[8]Using a similar approach, Peri (1998) finds that Mincerian returns to schooling are higher in more educated cities.
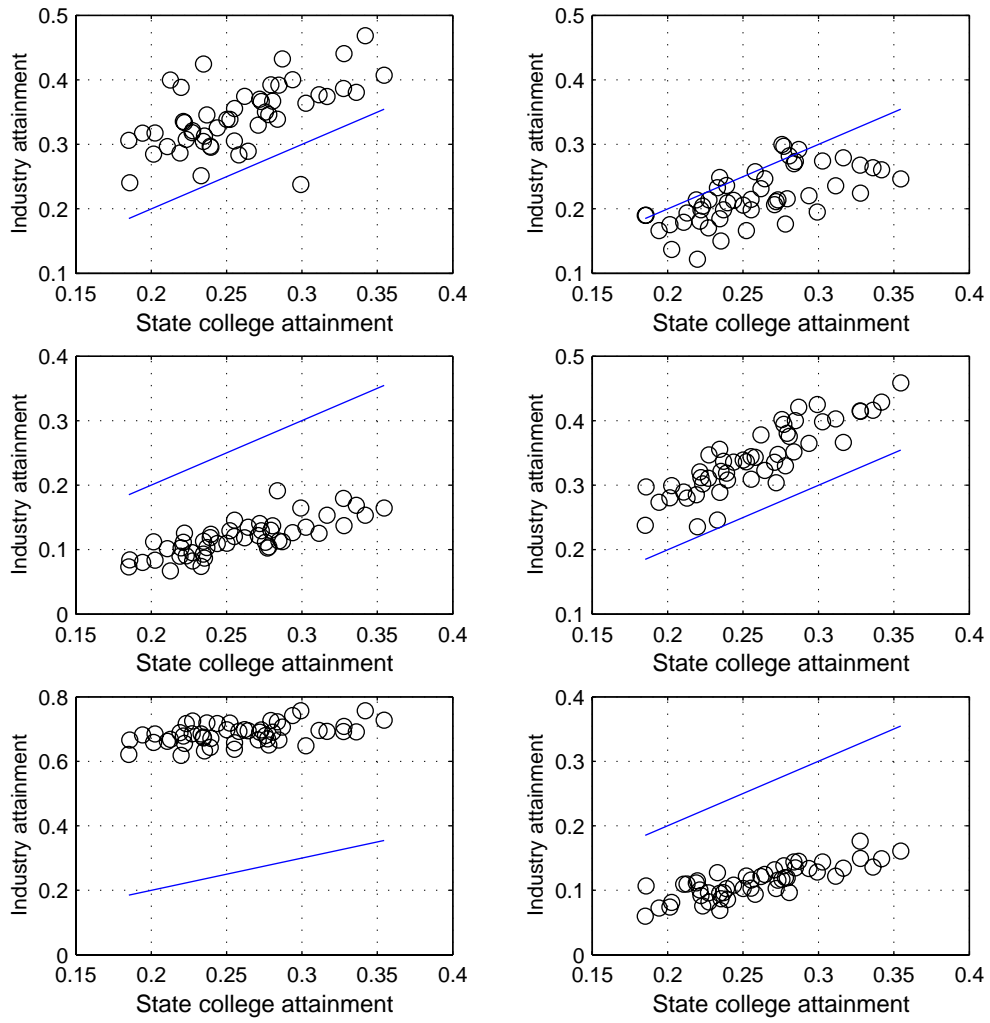
Figure 7: College attainment at the state and state-industry level.
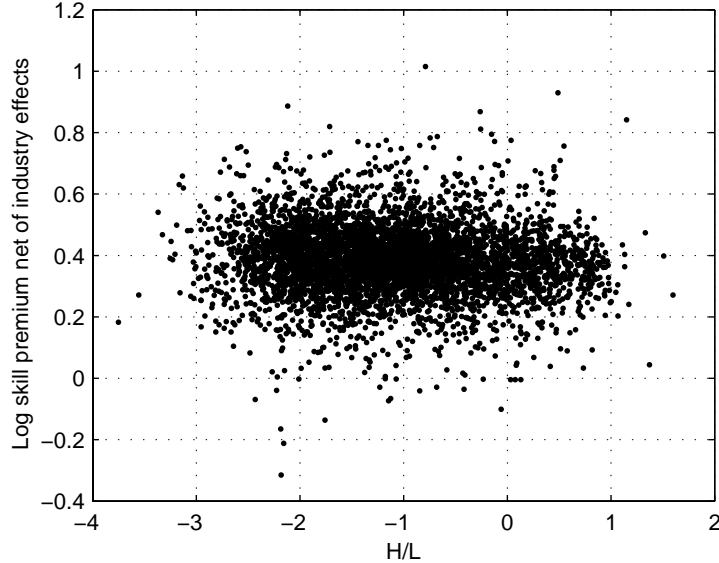
Figure 8: Skill premia for industry-state pairs

The first finding is that the skill premium equation, repeated here from (2),

$$\frac{w_{H,m,i}}{w_{L,m,i}} = \mu_i \; \varphi_m \; (H_{m,i}/L_{m,i})^{\rho-1}$$

predicts more than 90% of the variation in college attainment across state-industry pairs. That is, given estimates of μ_i, φ_m, and skill premia, (2) implies college attainment levels for each state-industry pair. Regressing predicted against observed attainment yields an R^2 greater than 0.9 in 1980 and 1990. This result supports the restriction imposed on the production functions that differences in φ_m are approximately *industry-neutral*. Whatever causes differences in the demand for skilled labor appears to affect all industries within a state symmetrically. This is an important finding that severely restricts the set of theories that could account for the data.

The second finding is that differences in skill premia are not important for understanding state attainment gaps. To support this claim, I predict state-industry attainment from (2) using U.S. average skill premia by industry. Figure 9 plots predicted against observed attainment using 1990 data. Regressing predicted against observed attainment yields a slope coefficient of 0.91 and an R^2 of 0.9. In other words, a model with perfect labor mobility and relative factor price equalization that correctly predicts the φ_m would account for 90% of the attainment variation across industries and states. This finding is consistent with Bernard et al.'s (2002) conclusion that relative factor price equalization holds across U.S. states, allowing for industry-neutral factor-specific technology differences.

To quantify the importance of skill bias differences, I aggregate the predicted values of $\eta_{m,i}$ to the state level using U.S. industry employment shares:

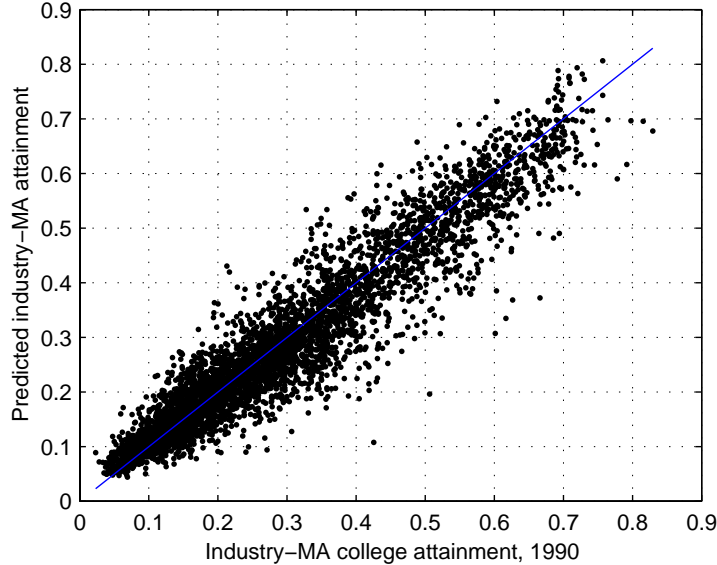$$\tilde{\eta}_m = \sum_i \tilde{\eta}_{m,i} \; \omega_{US,i} \tag{11}$$

15

Figure 9: Predicted attainment for industry-state pairs.

This removes the effects of industry specialization and retains the effects of differences in φ_m across states. Figure 10 plots the predicted against the observed state attainment for 1990 (all results are very similar for 1980). The slope of an OLS regression line is 0.47. In this sense, differences in φ_m account for nearly half of attainment differences. Note that this method imputes attainment for a rather large part of the work force in most states. The reason is that a given state typically has zero employment in many industries (Rosenthal and Strange 2002).[9] This raises concerns that small errors in the estimated φ_m could be enlarged by the imputation. If $\tilde{\eta}_m$ is calculated either with more aggregated industries or by only aggregating over industries where the state has positive employment, then the slope rises to around 0.8. However, this approach presumably fails to remove some effects of industry specialization.

To quantify the importance of industry specialization, I calculate the values of $\eta_{m,i}$ from (2) imposing the U.S. average φ. The predicted $\hat{\eta}_{m,i}$ are then aggregated to the state level using state employment weights:

$$\hat{\eta}_m = \sum_i \hat{\eta}_{m,i} \; \omega_{m,i} \tag{12}$$

This measure removes the effects of skill-bias differences, but retains the effects of industry specialization. Figure 11 plots predicted against observed state attainment. The slope is 0.53 leading me to conclude that specialization accounts for slightly more than half of attainment differences.

An alternative implementation uses observed rather than predicted attainment in (11) and (12). That is, the predicted $\tilde{\eta}_{m,i}$ are replaced with $\eta_{m,i}$ in (11) and the predicted $\hat{\eta}_{m,i}$

---

[9]Or a state has so few observations that industry attainment cannot be calculated reliably, in which case I drop the industry-state pair.
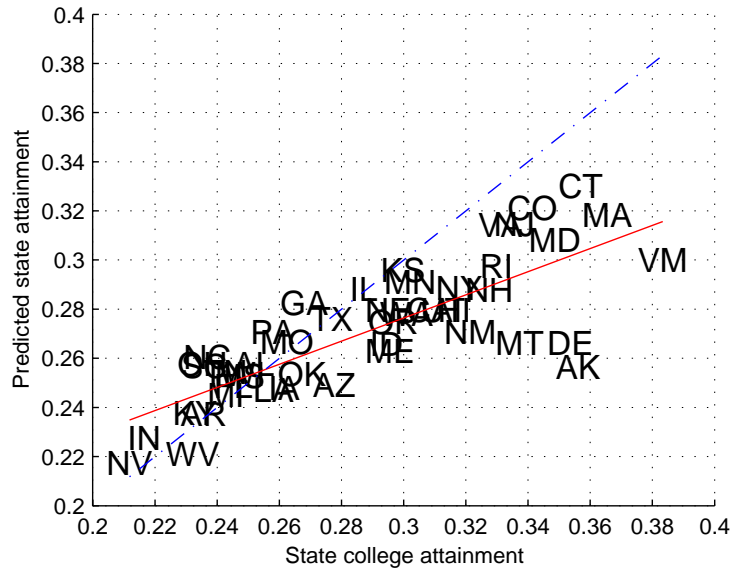
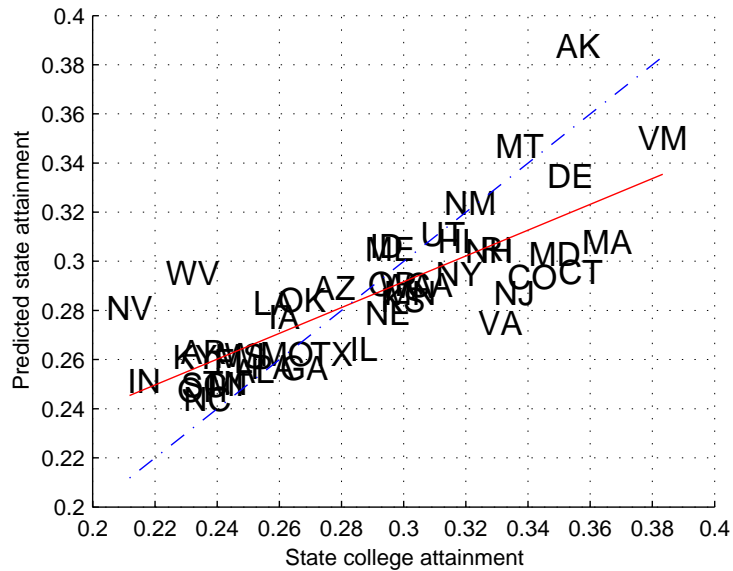Figure 10: State attainment net of industry specialization

Figure 11: State attainment net of skill bias differences

17

are replaced with U.S. average attainment for each industry in (12). This implementation attributes only about one-third of the observed attainment gaps to industry specialization. However, this method only permits to aggregate over industries where attainment is observed for each state in (11). This does not fully capture the effects of industry specialization.

Taken together, these observations suggest that industry specialization and differences in skill bias within industries account for roughly half of observed college attainment gaps across U.S. states. Differences in skill premia account for a negligible fraction.

5 Theories

This section outlines a model of educational attainment that qualitatively accounts for the facts documented in this paper:

1. Skill premia do not covary with attainment by location or by location-industry.

2. Educated places employ more educated labor in every industry.

3. Educated places specialize in skill intensive industries.

4. Educated places are endowed with skill-biased technologies across all industries (high φ_m).

5.1 A Reduced Form Model

It is useful to begin by describing a reduced form model which accounts for these observations. I only sketch this model because not all details are needed to derive the implications of interest. The basic idea is that differences in skill weights (φ_m) across locations lead to differences in attainment within industries and to industry specialization. The key model features are:

1. The world lasts for one period.

2. There are M locations and I industries.

3. The production functions are given by (1). Impose $A_{m,i} = A$ so that places differ only in their φ_m. Assume further that $\rho < 1$ based on empirical estimates.

4. All markets are perfectly competitive.

5. All goods are costlessly traded.

6. There are L unskilled workers and H skilled workers. Workers move costlessly between locations to maximize consumption.

7. Agglomeration costs (g) increase in employment: $g'(L_m + H_m) > 0$. A worker who earns w in location m consumes $w/g(L_m + H_m)$.

Such a theory accounts for the three observations highlighted above:

1. Labor mobility across industries equalizes industry wages in a given location: $w_{s,m,i} = w_{s,m} \ \forall \ i, \ s = H, L$. Mobility across locations equalizes wages net of migration costs, $w_{s,m} = w_s \ g\left(L_m + H_m\right)$, and implies relative factor price equalization. This may be a good description of the data. As pointed out by Glaeser and Mare (1999), wage rates divided by cost of living indices do not seem to vary systematically by city size, even though larger cities are more productive.

2. Free trade equalizes the prices of goods across locations: $p_{m,i} = p_i \ \forall \ m$. Therefore, by (2), places with high φ_m employ larger amounts of skilled labor in every industry: $\varphi_m > \varphi_{\hat{m}} \Rightarrow \eta_{m,i} > \eta_{\hat{m},i} \forall i$.

3. To see that places with high φ_m specialize in high μ industries, note that the unit cost of producing good i in place m is given by

$$c_{m,i} = w_{L,m} \ \left[1 + \left(\mu_i \ \varphi_m\right)^{1/(1-\rho)} \pi^{-\rho/(1-\rho)}\right]^{1-1/\rho} \tag{13}$$

where $\pi = w_{H,m}/w_{L,m}$ is the skill premium that is common to all places. It follows that places with high φ_m have a comparative advantage in products with high μ_i in the sense that

$$\varphi_m > \varphi_{\hat{m}} \wedge \mu_i > \mu_{\hat{i}} \Rightarrow \frac{c_{m,i}}{c_{m,\hat{i}}} < \frac{c_{\hat{m},i}}{c_{\hat{m},\hat{i}}} \tag{14}$$

To see that this is true, note that

$$\frac{\partial \ln c_{m,i}}{\partial \varphi_m} = - \left[\rho \varphi_m\right]^{-1} \left[1 + \left(\mu_i \ \varphi_m\right)^{1/(\rho-1)} \pi^{\rho/(1-\rho)}\right]^{-1} \tag{15}$$

It follows that places with high φ_m have an absolute advantage in all goods which is offset by higher wages. Since $|\partial \ln c_{m,i}/\partial \varphi_m|$ is increasing in μ_i, such places also have a comparative advantage in goods with high μ_i.

4. If places with high φ_m specialize in products with high μ_i and use more educated labor within each industry, then aggregate H/L is also high in these places.

It was shown earlier that the theory outlined here accounts for 90% of the variation in attainment across industry-state pairs, if it correctly predicts the φ_m (see figure 9). To what extent the model quantitatively accounts for industry specialization is an open question. The answer obviously depends on additional features of the model, such as household preferences.

The key issue for understanding why educational attainment differs across states is then to understand the determinants of φ_m. In the next section I argue that agglomeration theories are promising.[10] The reason is that larger and denser places employ more skilled workers without paying lower skill premia, suggesting that φ_m is related to the size or density of a market. This arises naturally in agglomeration theories. In a number of such theories, agglomeration economies arise because of the interaction of workers with specialized skills

---

[10]Theories of directed technical changed, such as the one proposed by Acemoglu (2002, 2003), may be a promising alternative. However, it is not clear how such theories would account for the observed relationship between education and city size or population density.

or firms with specialized knowledge. It appears natural to conjecture that the productivity of skilled workers benefits more from agglomeration economies than the productivity of unskilled workers. Below I present two examples of such theories based on matching of skilled workers and based on non-tradeable intermediate goods.

5.2 A Matching Model

In this section, I develop a model where the matching of workers to jobs generates a positive relationship between city population size and skill bias of the technology in use. The basic idea is that larger cities sustain more firms and hence permit better matches. The model is based on Duranton and Puga (2003).

The world lasts for one period. There are M cities, L unskilled workers, and H skilled workers. The timing of events is as follows:

1. Workers decide which city to live in so as to maximize expected consumption.

2. Skilled workers in each city draw skill types $\sigma_j \in [0,1]$ from a uniform distribution.

3. n_m firms enter the market in city m and pay a fixed cost c. Firms' types τ_i are evenly spaced along the unit circle.

4. Firms post skilled wage rates, $w_{H,m}^i$, and offer to hire any amount of skilled labor, $h_{m,i}$, at that wage.

5. Skilled workers decide which firm to work for. They earn $w_{H,m}^i$ net of a mismatch cost which depends on $|\sigma_j - \tau_i|$.

6. A Walrasian market for unskilled labor determines the wage rate, $w_{L,m}$, and firm employment, $l_{m,i}$.

7. Production of a single good takes place according to a version of (1):

$$Y_m^i = F(h_{m,i}, l_{m,i})$$
$$= \left[\mu \, (h_{m,i})^\rho + (l_{m,i})^\rho \right]^{1/\rho} \tag{16}$$

8. Output is consumed.

For simplicity, the model abstracts from multiple industries. Extending the argument to multiple industries along the lines of section 5.1 requires that agglomeration economies are not industry specific. That is, workers must be mobile across sectors so that working in a large city benefits all skilled workers symmetrically.

Workers: Workers choose the city and firm which offers the highest expected consumption level. A worker in city m who earns e consumes $e \, / \, g \, (L_m + H_m)$ where L_m is unskilled employment and H_m is skilled employment in city m. g represents a congestion cost, such as commuting, which is higher in larger cities. I assume $g' \geq 0$ and $g'' > 0$. Unskilled workers earn $w_{L,m}$ in city m. A skilled worker of type σ_j who works for the firm of type τ_i earns w_H^i net of a mismatch cost, $\chi \, |\sigma_j - \tau_i|$.

Firms: Unskilled labor is hired in a Walrasian market until the wage rate equals the marginal product:

$$w_{L,m} = F_l\left(h_{m,i}, l_{m,i}\right) \tag{17}$$

In the problem of setting the skilled wage offer, the firm takes as given the skilled wages offered by all other firms and hence the functional relationship between the wage offer and employment, $h_{m,i} = q_{m,i}\left(w_{H,m}^i\right)$. The wage offer is chosen to maximize profits:

$$\max_{l_{m,i}, w_{H,m}^i} \Pi_m^i = F\left(q_{m,i}\left(w_{H,m}^i\right), \, l_{m,i}\right) - w_{L,m}\, l_{m,i} - w_{H,m}^i\, q_{m,i}\left(w_{H,m}^i\right) \tag{18}$$

The first-order condition is

$$w_{H,m}^i = F_h(q_{m,i}\left(w_{H,m}^i, m, i\right), \, l_{m,i}) - q_{m,i}\left(w_{H,m}^i\right)/q_{m,i}'\left(w_{H,m}^i\right) \tag{19}$$

Firms pay less than marginal product because of monopsony power in the labor market. A firm that wants to enter the market pays a fixed cost c. Firms enter until profits are driven to zero:

$$F\left(h_{m,i}, l_{m,i}\right) - w_{L,m}\, l_{m,i} - w_{H,m}^i\, h_{m,i} = c \tag{20}$$

5.2.1 Symmetric equilibrium

Consider an equilibrium where all firms in location m offer the same skilled wage rate, $w_{H,m}$. Worker migration equalizes net expected earnings across locations. For unskilled workers this requires

$$w_L = w_{L,m} \, / \, g\left(L_m + H_m\right). \tag{21}$$

where w_L is the common real unskilled wage in all places. For skilled workers, the distance $|z| = |\sigma^j - \tau_i|$ is uniformly distributed on $[0, 0.5 \, n_m^{-1}]$. Expected skilled earnings are therefore given by $w_{H,m} - \chi/(4n_m)$. Migration then requires that expected real skilled earnings are equalized across places:

$$w_H = \frac{w_{H,m} - \frac{\chi}{4n_m}}{g\left(L_m + H_m\right)}. \tag{22}$$

In symmetric equilibrium the skilled labor supply function facing a firm satisfies

$$q_{m,i}(w_{H,m}^i)/q_{m,i}'\left(w_{H,m}^i\right) = \chi/n. \tag{23}$$

To see this, note that for any $w_{H,m}^i$ the firm hires workers up to the distance z where the worker is indifferent between firm i and its nearest competitor. This distance satisfies $w_{H,m}^i - \chi z = w_{H,m} - \chi\left(1/n_m - z\right)$ or

$$z = \frac{w_{H,m}^i - w_{H,m} + \chi/n_m}{2\chi} \tag{24}$$

The amount of skilled labor hired is then given by

$$\begin{aligned} q_{m,i}\left(w_{H,m}^i\right) &= 2\, z\, H_m \\ &= H_m\, \left[1/n_m + \left(w_{H,m}^i - w_{H,m}\right)/\chi\right] \end{aligned} \tag{25}$$

21

The marginal effect of the wage offer on employment is

$$q'_{m,i}\left(w^i_{H,m}\right) = H_m/\chi \tag{26}$$

which, together with $q_{m,i}\left(w^i_{H,m}\right) = H_m/n_m$, implies (23). Therefore the skilled wage rate in symmetric equilibrium is given by

$$w_{H,m} = F_h(H_m/n_m,\ L_m/n_m) - \chi/n_m \tag{27}$$

A *symmetric equilibrium* is then defined by a vector $(L_m, H_m, w_{L,m}, w_{H,m}, n_m)$ for each location and a pair (w_L, w_H) that satisfies:

- the firm's first order conditions (17) and (27);

- the free entry condition (20);

- the factor price equalization conditions (21) and (22);

- the labor market clearing conditions $\sum_m L_m = L$ and $\sum_m H_m = H$,

where $l_m = L_m/n_m$ and $h_m = H_m/n_m$. As is common in models with agglomeration economies, it is difficult to rule out multiple equilibria. Whether a unique equilibrium with a nondegenerate city size distribution exists depends on the functional form of the congestion cost, g.

5.2.2   Comparative statics

To derive the comparative statics properties of the model, I reduce the equilibrium conditions for a given location to 3 equations which determine n_m, η_m, and $L_m + H_m$. Relative factor price equalization implies that a high n_m is associated with a high η_m:

$$\mu\left(H_m/L_m\right)^{\rho-1} = \frac{w_H}{w_L} + \frac{\chi\left(1+1/4\right)}{n_m\ w_L\ g\left(L_m + H_m\right)} \tag{28}$$

The intuition is that more firms permit better matches, which raises the relative productivity of skilled labor. This is akin to a high φ_m in the reduced form model outlined above. To restore relative factor price equalization, it is necessary to employ more skilled labor, which reduces its marginal product.

The first-order condition for unskilled labor may be written as

$$F_l\left(H_m/n_m, L_m/n_m\right) = w_{Lm} = w_L\ g\left(L_m + H_m\right) \tag{29}$$

By virtue of constant returns to scale F_l is an increasing function of H_m/L_m. This implies that larger cities employ more skilled labor. The intuition is that larger cities must offer higher wages to unskilled workers in order to offset higher congestion costs.

The free entry condition (20) simplifies to

$$\eta_m = \frac{c\ n_m^2}{\chi\ \left(L_m + H_m\right)}. \tag{30}$$

To see this, note that constant returns to scale imply that

$$F\left(H_m/n_m, L_m/n_m\right) - F_l \, L_m/n_m - F_h \, H_m/n_m = 0$$

which implies

$$\chi/n_m \cdot \eta_m = c\, n_m \,/\, \left(L_m + H_m\right).$$

The main question of interest is: Do larger cities employ more skilled labor without featuring lower skill premia? The answer presumably depends on the reasons why cities of different sizes coexist. If all cities share the same parameters but there are multiple equilibria, it follows directly from (29) that larger cities employ more skilled workers. To see this, recall that F_l is an increasing function of η_m. Larger cities have higher congestion costs (g) and must therefore pay higher unskilled wages (F_l).

A natural alternative is that cities differ in their agglomeration costs. It does not matter exactly how g differs across cities as long as a non-degenerate city size distribution exists. Consider two cities that differ in their equilibrium population sizes $(L_m + H_m)$. The cities also differ in congestion costs. In the empirically interesting case, the larger city has higher congestion costs (g). This city must also employ more skilled labor (high η_m) to satisfy (29) and host more firms (high n_m) to satisfy (30). Finally, this is consistent with (28). Higher values of $L_m + H_m$ and n_m reduce the right-hand side, while a high η_m lowers the left-hand side of (28). This case offers a micro-foundation for larger cities employing more skill-biased (reduced form) technologies.

The alternative case where g is lower for larger cities could be an equilibrium outcome as well. Large cities would then pay lower absolute wages and have a fewer skilled workers by (29) than smaller cities. The comparison of η_m and n_m across cities would be indeterminate. While possible in principle, this case is not empirically interesting.

5.3   Skill-intensive Intermediate Inputs

Very similar implications can be derived from a model of non-traded, differentiated intermediate inputs. The basic idea is that larger cities offer more differentiated intermediate inputs. If the production of these inputs is skill-intensive (e.g., legal or consulting services), then the reduced form production function of a large city exhibits skill bias. The model presented here is based on Duranton and Puga (2003). As in the matching model, there are M locations, indexed by m, L unskilled workers and H skilled workers. Labor is perfectly mobile across locations. The final good is costlessly traded while intermediate goods are not traded.

Final goods: A single final good is produced by a representative firm in each location under perfect competition. The production function is

$$\begin{aligned} Y_m &= F\left(X_m, L_m\right) \\ &= \left[\mu X_m^\rho + L_m^\rho\right]^{1/\rho} \end{aligned}$$

where

$$X_m = \left[\int_0^{n_m} (x_j)^{\frac{1}{1+\varepsilon}} \, dj\right]^{1+\varepsilon}$$

is an aggregator of intermediate inputs, x_j, which are sold at prices q_j. The mass of intermediate inputs is n_m. The firm maximizes period profits:

$$\max F\left(X_m, L_m\right) - w_{L,m}L_m - \int_0^{n_m} q_j\ x_j\ dj.$$

The price of the final good is normalized to one. The first-order conditions equate factor rental prices with marginal products:

$$w_{L,m} = F_L\left(X_m, L_m\right) \tag{31}$$

$$q_j = F_X(X_m, L_m)\ (X/x_j)^{\frac{\varepsilon}{1+\varepsilon}} \tag{32}$$

The implied demand function for intermediate inputs is given by

$$x_j = X_m\ \left(\frac{q_j}{F_X}\right)^{-\frac{1+\varepsilon}{\varepsilon}} \tag{33}$$

Intermediate goods: Each intermediate input is produced by one firm. The production function uses skilled labor (h_j):

$$x_j = \beta\ h_j - \alpha$$

where $\alpha > 0$ is a fixed cost. Each firm maximizes profits in a monopolistically competitive market, taking the demand function (33) as given. Given the constant elasticity of demand, the optimal price is a fixed markup over marginal cost:

$$q_j = (1 + \varepsilon)\,\frac{w_{H,m}}{\beta} \tag{34}$$

where $w_{H,m}$ is the skilled wage rate in place m. Intermediate goods producers costlessly enter and exit the market. Free entry drives profits to zero at each point in time:

$$q_j x_j - w_{H,m} h_j = 0 \tag{35}$$

5.3.1 Equilibrium

The equilibrium quantity produced of each intermediate good implied by (35) is

$$x_j = \frac{\alpha}{\varepsilon} \tag{36}$$

which implies that each firm hires

$$h = h_j = \frac{\alpha}{\beta}\left(1 + 1/\varepsilon\right)$$

skilled workers. The number of intermediate firms entering must clear the market for skilled labor:

$$n_m = \frac{H_m}{h} = C_1 H_m$$

where $C_1 = \frac{\beta}{\alpha}\frac{\varepsilon}{1+\varepsilon}$. Therefore $X_m/x_j = n_m^{1+\varepsilon} = C_1^{1+\varepsilon} H_m^{1+\varepsilon}$. The equilibrium price q then follows from the demand function (32):

$$q_m = F_X C_1^\varepsilon H_m^\varepsilon \tag{37}$$

The skill premium is given by

$$\frac{w_{H,m}}{w_{L,m}} = \frac{\beta}{1+\varepsilon} C_1^\varepsilon H_m^\varepsilon \frac{F_X}{F_L}$$

$$= \frac{\beta}{1+\varepsilon} C_1^\varepsilon H_m^\varepsilon \mu \ (L_m/X_m)^{1-\rho} \tag{38}$$

$$= C_2 \mu \ (L_m/H_m)^{1-\rho} H_m^{\rho\varepsilon} \tag{39}$$

where $C_2 > 0$ is a function of parameters. The reduced form production function of city m may then be written as

$$Y = [\mu_m H_m^\rho + L_m^\rho]^{1/\rho} \tag{40}$$

where the skill bias parameter

$$\mu_m = \mu \ C_2 \ H_m^{\varepsilon\rho} \tag{41}$$

is increasing in the city's skilled population. Note that the expressions for the production function (40) and the skill premium (39) are the same that would be obtained from a model where more skilled places employ more skill-biased technologies (as described by (41)), perhaps due to a local human capital externality.

Equilibrium with migration: To complete the model and derive its implications for cities of different sizes, I add a migration block with congestion costs which is identical to that of the matching model. The timing of events is as follows:

1. Workers choose where to live and work.

2. Intermediate goods producers hire skilled labor and produce.

3. Final goods producers hire unskilled labor and purchase intermediates.

4. Final goods output is consumed.

An equilibrium with migration consists of a labor allocation (L_m, H_m) and prices $(w_{H,m}, w_{L,m}, w_L, w_H)$ that satisfy:

1. The skill premium equation (39).

2. The first-order condition for unskilled labor (31).

3. The factor price equalization conditions

$$w_{H,m} = w_H \ g \left(L_m + H_m\right) \tag{42}$$
$$w_{L,m} = w_L \ g \left(L_m + H_m\right) \tag{43}$$

4. Market clearing: $\sum_m L_m = L$ and $\sum_m H_m = H$.

5.3.2 Comparative statics:

As in the matching model, there are two reasons why cities of different sizes can coexist: multiple equilibria and differences in congestion costs. In both cases, the model delivers essentially the same predictions as the matching model.

Consider first the case of multiple equilibria. Larger cities then have higher congestion costs. To attract workers, they need to pay higher wages. For unskilled workers, higher wages require larger X_m/L_m ratios by (31). To satisfy relative factor price equalization, by (39) larger cities must have larger fractions of skilled workers. Finally, the free entry condition implies that larger cities have more varieties of intermediate goods (larger n_m).

Alternatively, consider the case where the equilibrium is unique, but cities differ in their congestion costs, g. In the empirically interesting case, larger cities have higher congestion costs. The implications are then the same as in the case of multiple equilibria.

The main point that emerges from these examples is that matching skilled workers to jobs and differentiated intermediate goods offer natural micro-foundations for the reduced form model of section 5.1. In both cases larger cities employ more skilled-biased (reduced form) technologies and employ more skilled workers without paying lower skill premia than do smaller cities.

The more general point is: The micro-foundations of agglomeration models suggest that skilled workers benefit more from agglomeration than do unskilled workers. This may be because engineers benefit from interaction with other engineers while janitors do not. Or it may be because firms use non-tradable business services that require skilled labor, such as accounting or consulting services. In either case, agglomeration economies lead to skill-bias in reduced form technologies and offer a possible explanation for the observations documented in this paper.

6   Conclusion

The fraction of persons holding a college degree differs two-fold across U.S. states. This paper documents the relationship between educational attainment, skill premia, and industry specialization for U.S. cities and states. The empirical findings suggest that attainment differences stem from local variations in the skill-bias of technologies in use. Impediments to labor mobility or other sources of regional wage differences appear to play a minor role. Models of agglomeration are developed that account for the observed relationship between attainment, skill premia, and industry specialization.

Future research should attempt to measure the fraction of educational attainment differences across cities and states that are due to agglomeration economies and other factors.

Appendix

A   Data Sources and Variable Construction

Census Data: The main data sources are the Decennial Censuses from 1940 to 1990. For 1980 and 1990, the 5% State Samples are used. Data for earlier years are obtained from IPUMS. The unit of observation is the individual. Table 2 shows summary statistics at the state level for 1990. The number of persons observed in each state exceeds 10,000. Samples for earlier years are smaller. In 1940, all states but Nevada have more than 1,000 observations.

Individual wage rates are calculated as the ratio of labor income to hours worked per year. Wage rates below 10% of the median and above 100 times the median are deleted as likely measurement errors.

MA codes are taken from the variable MSAPMSA for 1990 and from SMSA in 1980. To determine whether a person lives in an MA in 1990, I check whether the last 3 digits of MSAPMSA equal 999. In 1980, I check whether AREATYPE equals 5 ("outside smsas").

Other Data Sources: The population density of states is calculated from county level data. Population size and land area are taken from the 1998 edition of the Census Bureau's *USA Counties* database. The calculation of Ciccone and Hall's (1996) measure of density requires data on state gross product. These are taken from the regional accounts of the Bureau of Economic Analysis (http://www.bea.gov/bea/regional/gsp/) for the years 1979 and 1989.

State	N	College	η	π	Frac.MA
AL	86282	0.21	0.22	1.61	0.78
AK	14135	0.26	0.28	1.37	0.44
AZ	78219	0.24	0.26	1.56	0.81
AR	48696	0.18	0.19	1.55	0.52
CA	671010	0.27	0.30	1.50	0.96
CO	82582	0.32	0.33	1.54	0.84
CT	75613	0.35	0.34	1.50	1.00
DE	14759	0.27	0.27	1.56	0.68
FL	288392	0.23	0.24	1.59	0.95
GA	152247	0.24	0.26	1.57	0.82
HI	25384	0.28	0.29	1.38	0.75
ID	21277	0.22	0.23	1.53	0.23
IL	262835	0.27	0.28	1.51	0.89
IN	124333	0.20	0.20	1.47	0.80
IA	58469	0.22	0.22	1.51	0.75
KS	53789	0.27	0.27	1.53	0.69
KY	80475	0.18	0.20	1.51	0.59
LA	89009	0.20	0.24	1.56	0.80
ME	28855	0.24	0.24	1.46	0.65
MD	118035	0.32	0.34	1.50	0.97
MA	143355	0.35	0.35	1.47	0.97
MI	205977	0.21	0.23	1.48	0.83
MN	101648	0.27	0.28	1.52	0.72
MS	54540	0.19	0.21	1.55	0.41
MO	111719	0.24	0.24	1.52	0.72

Table 2: Summary statistics for U.S. states, 1990.

Notes: N: number of persons in the sample. College: fraction of persons holding a college degree. η: college attainment in efficiency units. π: skill premium. Frac.MA: fraction of population residing in metropolitan areas.

State	N	College	η	π	Frac.MA
MT	17350	0.25	0.26	1.41	0.29
NE	33520	0.25	0.25	1.54	0.56
NV	29235	0.18	0.19	1.48	0.91
NH	27570	0.30	0.29	1.49	0.74
NJ	191091	0.34	0.33	1.52	1.00
NM	34710	0.25	0.28	1.56	0.51
NY	417839	0.30	0.31	1.54	0.95
NC	154074	0.22	0.22	1.60	0.65
ND	12920	0.26	0.26	1.54	0.83
OH	242169	0.21	0.23	1.52	0.85
OK	67750	0.23	0.24	1.54	0.84
OR	64259	0.26	0.26	1.47	0.70
PA	267423	0.25	0.25	1.56	0.90
RI	22062	0.28	0.28	1.51	1.00
SC	74539	0.20	0.21	1.58	0.64
SD	13878	0.23	0.23	1.50	0.44
TN	111636	0.21	0.22	1.58	0.74
TX	395277	0.25	0.27	1.62	0.88
UT	34620	0.26	0.28	1.50	0.81
VM	13331	0.30	0.30	1.44	0.44
VA	141962	0.29	0.32	1.59	0.84
WA	113650	0.28	0.28	1.44	0.83
WV	38888	0.16	0.19	1.42	0.54
WI	109786	0.23	0.23	1.50	0.75
WY	10461	0.22	0.22	1.42	0.51
Mean	112633	0.25	0.26	1.52	0.73
Min	10461	0.16	0.19	1.37	0.23
Max	671010	0.35	0.35	1.62	1.00

Table 2: Summary statistics for U.S. states, 1990 (cont.).

B   Factor Quality Differences

The analysis in the main text abstracts from the possibility that unmeasured skills differ across locations. Yet the data are consistent with an interpretation where more educated places employ workers with more unmeasured skills.

To capture this possibility, assume that each skilled (unskilled) person in location m is endowed with $\alpha_{H,m}$ $(\alpha_{L,m})$ efficiency units of skilled (unskilled) labor. Observed factor inputs are then $\tilde{H}_m = H_m/\alpha_{H,m}$ and $\tilde{L}_m = L_m/\alpha_{L,m}$. Since the empirical analysis uses ratios of skilled to unskilled labor inputs and wages, the findings are affected if the ratio $\alpha_{H,m}/\alpha_{L,m}$ differs across locations.

Consider first the case where labor mobility or trade in goods equalizes relative factor prices across locations. With identical technologies (φ_m), (2) implies that H_m/L_m is equalized across locations. Therefore, the wage bill ratio in (8) should be the same in all locations. This prediction is overwhelmingly rejected in the data where the wage bill ratio is very strongly

correlated with college attainment. Bernard and Schott (2002) use a similar logic to test for factor price equalization across U.S. states.

However, if relative factor price equalization is violated, the model becomes vacuous. For any data on relative skill prices $(\tilde{w}_H/\tilde{w}_L)$ and quantities $\left(\tilde{H}/\tilde{L}\right)$ there exists a ratio $\alpha_{H,m}/\alpha_{L,m}$ such that the data are consistent with cost minimization. Differences in unmeasured skills cannot be distinguished observationally from differences in skill bias across locations.

The main reason why the possibility of unmeasured skill differences is not explored further in this paper is parsimony. Observed skills do not differ systematically with college attainment. Theories where labor mobility or trade in goods equalize relative factor prices offer a parsimonious explanation for this observation. By contrast, theories of unmeasured skills require a selection mechanism which ensures that locations with large skill quantities $\left(\tilde{H}/\tilde{L}\right)$ also have high relative skill qualities (α_H/α_L) but low skill premia in efficiency units (assuming that $\rho < 1$). Moreover, the quantitative relationship between $\tilde{H}/\tilde{L}$ and α_H/α_L must be such that states and MAs with different attainment levels have approximately the same measured skill premia. The fact that more educated states employ larger amounts of educated labor in all industries imposes additional requirements on this selection process. In this paper, I therefore focus on models with perfect labor mobility and relative factor price equalization.

References

[1] Abdel-Rahman, Hesham M.; Masahisa Fujita (1990). "Product variety, Marshallian externalities, and city sizes." *Journal of Regional Science* 30(2):165-183.

[2] Acemoglu, Daron (2002). "Directed technical change." *Review of Economic Studies* 69: 781-810.

[3] Acemoglu, Daron (2003). "Patterns of skill premia." *Review of Economic Studies* 70: 199-230.

[4] Baier, Scott; Sean Mulholland; Robert Tamura; Chad Tuner (2004). "Income and education of the states of the United States: 1840-2000." Mimeo. Clemson University.

[5] Bernard, Andrew B.; Peter K. Schott (2002). "Factor price equality and the economies of the united States." Mimeo. Yale University.

[6] Black, Duncan; Vernon Henderson (1999). "A theory of urban growth." *Journal of Political Economy* 107(2): 252-84.

[7] Bound, John; Jeffrey Groen; G.Gabor Kezdi; Sarah Turner (2004). "Trade in university training: cross-state variation in the production and stock of college-educated labor." *Journal of Econometrics* 121(1-2): 143-173.

[8] Ciccone, Antonio; Giovanni Peri (2004). "Long-run substitutability between more and less educated workers: evidence from U.S. states 1950-1990." Mimeo. University of California, Davis.

[9] Duranton, Gilles; Diego Puga (2003). "Microfoundations of urban agglomeration economies." Mimeo. London School of Economics.

[10] Eaton, Jonathan; Zvi Eckstein (1997). "Cities and growth: theory and evidence from France and Japan." *Regional Science and Urban Economics* 27: 443-74.

[11] Glaeser, Edward (1999). "Learning in cities." *Journal of Urban Economics* 46: 254-277.

[12] Glaeser, Edward; D. Mare (2001). "Cities and Skills," *Journal of Labor Economics* 19(2): 316-342.

[13] Goldin, Claudia (1998). "America's graduation from high school: the evolution and spread of secondary schooling in the twentieth century." *Journal of Economic History* 58: 345-74.

[14] Goldin, Claudia (1999). "Egalitarianism and the returns to education during the great transformation of American education." *Journal of Political Economy* 107(6 part 2) S65-94.

[15] Helsley, Robert W.; William C. Strange (1990). "Matching and agglomeration economies in a system of cities." *Regional Science and Urban Economics* 20(2):189-212.

[16] Kaboski, Joseph P. (2003). "Endogenous Mincerian Returns: Explaining Cross-Country Variation in the Returns to Schooling." Mimeo. Ohio State University.

[17] Krusell, Per; Lee E. Ohanian; Jose-Victor Rios-Rull; Giovanni Violante (2000). "Capital-skill complementarity and inequality: a macroeconomic analysis." *Econometrica* 68(5): 1029-53.

[18] Peri, Giovanni (1998). "Human capital and productivity in U.S. cities." Mimeo. UC Davis.

[19] Rauch, James, E. (1993). "Productivity gains from geographic concentration of human capital: Evidence from the cities." *Journal of Urban Economics* 34: 380-400.

[20] Rosenthal, Stuart S.; William C. Strange (2002). "Evidence on the nature and sources

of agglomeration economies." Mimeo. Syracuse University.

.