Educational Attainment in U.S. Cities

Lutz Hendricks
Iowa State University, Department of Economics
CESifo, Munich; CFS, Frankfurt
Preliminary. January 25, 2007

Abstract

This paper proposes a theory of educational attainment differences across U.S. metropolitan areas. The theory is motivated by the finding that employment in business services predicts more than 70% of the observed cross-city variation in education. In the model, agglomeration economies in the production of business services, which are complementary with skilled labor, account for cross-city variation in education. The theory makes a number of testable predictions which find strong support in U.S. data.

Key words: Education, human capital, cities, agglomeration economies.

JEL: J24

1 Introduction

Educational attainment differs dramatically across U.S. metropolitan areas ("cities"). In the year 2000, more than half of the adult population held a college degree in the five most educated cities, compared with only 10% in the five least educated cities (see section 4.1 for details on how these figures are constructed). The purpose of this paper is to propose a theory of educational attainment differences across cities.

The motivation for the proposed theory is shown in figure 1. The vertical axis shows educational attainment for 297 cities in the year 2000. Attainment is defined as the ratio of skilled (college degree) to unskilled (less than college degree) labor input. The horizontal axis shows a statistic, ϕ_m, that is derived solely from data on employment in *business services*. The point of figure 1 is: knowing each city's employment in business services is sufficient to predict 75% of the variation in attainment across cities.[1]

Figure 1 is striking because the construction of ϕ_m does not use any information about a number of city characteristics that might plausibly be related to education. These include city size (suggesting agglomeration economies), industry composition, skill prices (suggesting differences in skill supplies), or geographic location.

Theory. Motivated by figure 1, the main idea of the proposed theory is that business services complement skilled labor in the production of final goods. Section 2 describes the model in detail. The world is endowed with fixed quantities of skilled and unskilled labor. Workers move between cities and industries to maximize their earnings, net of congestion costs. Business services are non-traded intermediate inputs in the production of final goods, which are costlessly traded between cities. The production of business services is subject to agglomeration economies.

---

[1]That is, the R^2 obtained from regressing the logarithm of educational attainment on the logarithm of $\varphi_{m,x}$ is 0.75.
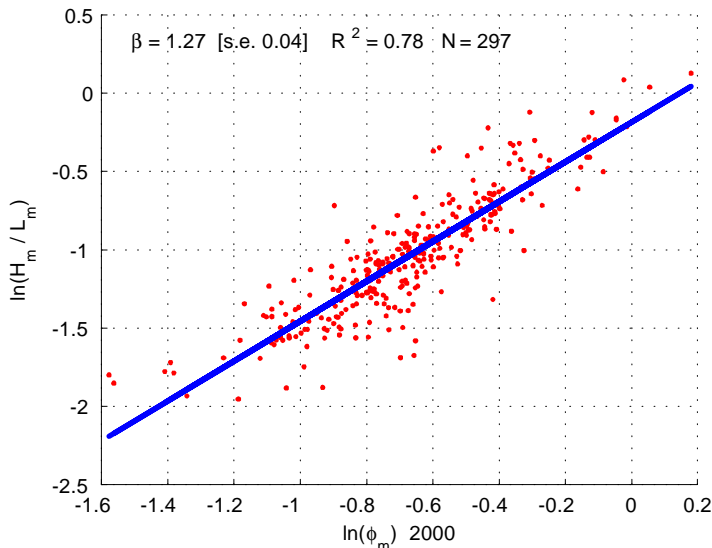
Figure 1: Attainment and skill bias

Section 3 derives a number of testable model predictions. In equilibrium, cities of different sizes coexist. This is due to the opposing forces of congestion costs and agglomeration economies. The model's main prediction is that *cities with larger business services sectors are more educated*. There are two reasons for this. (i) Cities with large services sectors *specialize* in skill intensive industries. (ii) Even though labor mobility equalizes skill premia across locations, cities with larger services sectors employ more skilled labor *within* each industry. Thus, to an econometrician who does not observe business services inputs, cities with larger services sectors appear to employ more skill biased technologies in all industries. Finally, due to agglomeration economies, cities with larger services sectors are more productive and pay higher wages.

The intuition for these predictions is as follows. Agglomeration economies reduce the price of services in cities with larger services sectors. This raises the demand for skilled labor in two ways. (i) Since services complement skilled labor in production, the demand for skilled labor increases within each industry. (ii) Cities with cheaper services also have a comparative advantage in skill-intensive final goods and specialize in their production.

Empirical evaluation. To evaluate the model's predictions, section 4 constructs a dataset spanning 297 cities and 126 industries. The main data sources are the 1980 to 2000 U.S. Decennial Censuses. All of the model's predictions find strong support in the data. Notably, more educated cities employ more skilled labor in each industry, even though their skill premia do not differ systematically from less educated cities. More educated cities also employ larger fractions of their labor in business services. This is consistent with the proposed hypothesis that scale economies in business services affect the demand for skilled labor across industries in a city.

The model makes quantitative predictions about educational attainment for cities and industries. It accounts for nearly 80% of the observed variation in education across city-industry cells and for roughly half of the observed cross-city differences in education *within industries*.

Consistent with the model's prediction, more educated cities specialize in skill-intensive industries. However, industry specialization accounts for only one-quarter of cross-city attainment gaps. The remaining three-quarters are due to within-industry differences in education across cities, which

the model attributes to variations in services prices.

Related literature.   To my knowledge, this is the first quantitative theory of educational attainment differences across U.S. cities. However, a number of previous papers have studied closely related questions.

Eaton & Eckstein (1997) and Glaeser (1999) develop models where cities of different human capital levels coexist. This is due to the interaction of congestion costs and human capital spillovers. In my model, cities are not places of learning. Instead, cities with large business services sectors are highly educated because they attract skilled workers from other cities.

Several papers argue that technology adoption is central for understanding the *changes* in cities' education over time. In Berry & Glaeser (2005), a city's initial endowment of skilled labor determines firms' investment in skill biased technology adoption. Thus, cities' education levels diverge over time. Lewis (2004, 2005) and Beaudry, Doms & Lewis (2005) study how cities absorb changes in the supply of skilled labor due to immigration. They argue that exogenous increases in skilled labor endowments induce skill biased technology adoption. Labor inflows are absorbed not by changing a city's industry mix, but by changing the ratio of skilled to unskilled employment within industries.

Peri (1998) observes that more educated cities pay higher skill premia. He interprets this as evidence of skill biased technology adoption in cities with abundant skilled labor.

2   Model

The model is built around two central ideas: (i) the production of business services is subject to agglomeration economies and (ii) business services complement skilled labor in the production of final goods.

The world lasts for one period. There are M cities, indexed by m, and I industries, indexed by i. $I-1$ industries produce final goods, which are traded costlessly between cities. Industry x produces a nontraded intermediate good: business services. Agglomeration economies imply increasing returns to scale in the production of services. All industries are perfectly competitive. The economy is endowed with H skilled workers and with L unskilled workers. These move costlessly between cities and industries to maximize their labor earnings.

Notation.   It is useful to fix notation before describing the model details. Skilled employment in city i and industry m is denoted by $H_{m,i}$. Unskilled employment is $L_{m,i}$. Total skilled employment in city m is given by $H_m = \sum_i H_{m,i}$. Unskilled employment is defined analogously as $L_m = \sum_i L_{m,i}$. My measure of educational attainment is the ratio of skilled to unskilled labor input:

$$\lambda_{m,i} = \ln\left(H_{m,i}/L_{m,i}\right) \tag{1}$$

Free trade equalizes the prices of the $I-1$ traded goods, denoted by p_i. The price of services ($p_{m,x}$) differs across cities. Labor mobility equalizes wages across (active) industries within a city. Skilled labor earns $w_{H,m}$ and unskilled labor earns $w_{L,m}$. The skill premium is defined as $\nu_m = w_{H,m}/w_{L,m}$.

2.1   Workers

Workers are either skilled or unskilled. The representative skilled worker inelastically supplies H units of labor. She derives utility from consuming the $I-1$ final goods and chooses consumption levels $C_{H,i}$ and labor supplies H_m to solve

$$\max U\left(C_{H,1}, ..., C_{H,I-1}\right) \tag{2}$$

3

subject to the budget constraint

$$\sum_i p_i C_{H,i} = \sum_m \frac{w_{H,m} H_m}{\chi_m} \tag{3}$$

and the time constraint

$$H = \sum_m H_m \tag{4}$$

The worker divides the time endowment H between the M cities so as to maximize earnings net of an agglomeration cost χ_m. The worker takes χ_m as given. However, in equilibrium, the agglomeration cost depends on the size of the city:

$$\chi_m = \Gamma_m \left(H_m + L_m \right)$$

with $\Gamma'_m > 0$. Evidently, labor mobility equates wages, net of agglomeration costs, across all (active) cities:

$$w_{H,m} = w_H \, \Gamma_m \left(H_m + L_m \right) \tag{5}$$

A solution to the skilled workers problem yields demand functions $D_{H,i} \left(p_1, ..., p_{I-1}; w_H \right)$ for final goods. In addition, labor supplies (H_m) must satisfy the time constraint (4) and the no arbitrage condition (5). The representative unskilled worker solves an analogous problem, which yields demand functions $D_{L,i} \left(p_1, ..., p_{I-1}; w_L \right)$ and labor supplies L_m that satisfy the time constraint

$$L = \sum_m L_m \tag{6}$$

2.2   Final goods

The production function for industry i is given by

$$Y_{m,i} = F_i \left(G \left(H_{m,i}, X_{m,i} \right), L_{m,i} \right) \tag{7}$$

where $Y_{m,i}$ is output and $X_{m,i}$ denotes purchased business services. Both F_i and G have constant returns to scale. F_i is common to all cities, but differs by industry. G is common across industries and cities. Firms maximize profits

$$\max p_{m,i} Y_{m,i} - w_{L,m} L_{m,i} - w_{H,m} H_{m,i} - p_{m,x} X_{m,i} \tag{8}$$

The first-order conditions equate marginal products to factor prices:

$$w_{L,m}/p_{m,i} = F_{iL} \left(G_{m,i}/L_{m,i} \right) \tag{9}$$

$$w_{H,m}/p_{m,i} = F_{iG} \left(G_{m,i}/L_{m,i} \right) G_H \left(X_{m,i}/H_{m,i} \right) \tag{10}$$

$$p_{m,x}/p_{m,i} = F_{iG} \left(G_{m,i}/L_{m,i} \right) G_X \left(X_{m,i}/H_{m,i} \right) \tag{11}$$

where $G_{m,i} = G \left(H_{m,i}, X_{m,i} \right)$. The partial derivatives, $F_{iz} = \partial F_i / \partial z$, depend on input ratios due to constant returns to scale. A solution to the firm's problem consists of $Y_{m,i}, L_{m,i}, H_{m,i}, X_{m,i}$ that satisfy the firm's first-order conditions and the production function. For the purpose of taking the model to the data I assume the F_i has a constant elasticity of substitution, $(1 - \rho)^{-1}$:

$$F_i = \left[\mu_i G \left(X_{m,i}, H_{m,i} \right)^\rho + (1 - \mu_i) L_{m,i}^\rho \right]^{1/\rho} \tag{12}$$

For reasons discussed in section 4.1, I assume $\rho > 0$.

4

2.3 Business services

The production function for business services is given by

$$Y_{m,x} = \Xi_m F_x \left(G \left(X_{m,x}, H_{m,x} \right), L_{m,x} \right) \tag{13}$$

This differs from the production function for final goods only in the productivity term Ξ_m. F_x exhibits constant returns to scale. Services firms maximize

$$\max p_{m,x} \Xi_m F_x \left(G \left(X_{m,x}, H_{m,x} \right), L_{m,x} \right) - w_{H,m} H_{m,x} - w_{L,m} L_{m,x} - p_{m,x} X_{m,x} \tag{14}$$

The first-order conditions are the same as for final goods producers, except for the Ξ_m factors. A solution to the business services firm's problem consists of $Y_{m,x}, L_{m,x}, H_{m,x}, X_{m,x}$ that satisfy the firm's first order conditions and the production function.

A key feature of the model is *agglomeration economies* in the production of business services. These are modeled as an externality. While firms take Ξ_m as given, in equilibrium it depends on the scale of the services sector:

$$\Xi_m = F_x \left(G_{m,x}, L_{m,x} \right)^\varepsilon, \quad \varepsilon > 0 \tag{15}$$

2.4 Competitive equilibrium

A competitive equilibrium consists of an allocation $(C_{H,i}, C_{L,i}, L_{m,i}, H_{m,i}, X_{m,i}, Y_{m,i}, H_m, L_m)$ and a price system $(w_{H,m}, w_{L,m}, w_H, w_L, p_{m,x}, p_i)$ such that

1. Workers maximize utility, given prices and agglomeration costs.

2. Final goods and services producers maximize profits, given prices.

3. Markets clear.

Labor market clearing is implied by the household's time constraints (4) and (6). The markets for final goods clear when

$$\sum_m Y_{m,i} = D_{H,i} \left(p_1, ..., p_{I-1}, w_H \right) + D_{L,i} \left(p_1, ..., p_{I-1}, w_L \right), \quad i \neq x \tag{16}$$

The market for business services clears when

$$Y_{m,x} = \sum_i X_{m,i} \quad \forall m \tag{17}$$

2.5 Discussion

The idea that business services are a source of agglomeration economies is not new. Duranton and Puga (2003) develop a model of differentiated business services that exhibits scale economies. The novel idea of this paper is to link business services to the demand for skilled labor in a city.

Most of the model's implications derived below do not depend on the details of the demand functions for final goods. The demand specification is therefore kept simple. Since final goods are not differentiated across cities, the model makes extreme predictions about industry specialization (see Proposition 6). In the empirical evaluation of the model, I therefore focus on its implications for attainment in city-industry pairs, rather than for city attainment.

The assumption of perfect labor mobility is strong. It is responsible for equalizing skill premia across cities. To check whether this assumption is reasonable, section 4.8 shows that the correlation between skill premia and attainment is very weak across cities and across city-industry pairs.

3   Empirical Implications

This section derives empirically testable implications of the model. I focus on equilibria where multiple cities coexist. First, I show that cities differ in service prices,

$$\pi_m \equiv p_{m,x}/w_{H,m} \tag{18}$$

Then I characterize how allocations and prices differ between cities with high and low service prices. I find that city attainment is negatively related to π_m. Section 4 takes the model implications to the data and shows that cities with low services prices have the properties of highly educated cities in the data.

Preliminary results.   It is easy to show that firms in low π_m cities are characterized by high $X_{m,i}/H_{m,i}$, $G_{m,i}/L_{m,i}$, $G_{m,i}/H_{m,i}$, and G_H, as well as by low G_X, and F_{iG}/F_{iL}. In addition, the final goods firm's first order conditions imply that $\pi_m = \frac{G_X(X_{m,i}/H_{m,i})}{G_H(X_{m,i}/H_{m,i})}$, so that the ratio of services to skilled labor inputs is determined by a common function across all cities and industries:

$$X_{m,i}/H_{m,i} = \Psi\left(\pi_m\right) \tag{19}$$

with $\Psi' < 0$. The intuition is that cheap services lead firms to substitute services for skilled labor. Since G is common to all industries and cities, the optimal input ratio $X_{m,i}/H_{m,i}$ only depends on the relative price of the two factors.

An additional assumption is required to ensure that firms facing cheap services employ larger amounts of skilled relative to unskilled labor. Define $g\left(X/H\right) \equiv G\left(X, H\right)/H$.

Assumption A1: $G_H\left(X/H\right) g\left(X/H\right)^{\rho-1}$ is increasing in X/H.

This assumption is essential for most of the paper's results. It formalizes the idea that skilled labor and services are sufficiently complementary, so that lower services prices (π_m) lead firms to employ more skilled labor. To see why A1 is needed, consider how a firm reacts to a lower π_m. Substitution towards cheaper services implies that firms hire more of them: $G_{m,i}/H_{m,i}$ and $G_{m,i}/L_{m,i}$ both increase. Whether $H_{m,i}/L_{m,i}$ increases depends on how substitutable factor inputs are in production.

If H is easily substituted for X, firms respond to cheaper services by "outsourcing" their skilled labor, i.e., $X_{m,i}$ increases and $H_{m,i}$ declines so that $H_{m,i}/L_{m,i}$ may decline. Intuitively, firms replace their own accountants with hired consultants. However, if X and H are strong complements, firms will increase both $X_{m,i}$ and $H_{m,i}$ in response to cheaper services. In that case $H_{m,i}/L_{m,i}$ increases. Rather than replacing their accountants, these firms hire additional accountants to work with external consultants. Assumption A1 ensures that firms respond in this second way.

A1 is not highly restrictive. For example, A1 holds if $G\left(X, H\right) = X^\alpha H^{1-\alpha}$ and $\rho > 0$. In this case,

$$g\left(X/H\right)^{\rho-1} G_H\left(X/H\right) = [X/H]^{\alpha(\rho-1)} \ [X/H]^\alpha$$

is increasing in X/H.

3.1   Coexistence of cities with different prices

The focus of the analysis is the comparison of cities with different services prices, which correspond (by Corollary 7 below) to different attainment levels. Proposition 1 shows that, in equilibrium, cities with different services prices coexist. This requires a technical assumption which rules out equilibria where cities of different sizes happen to have the same agglomeration costs.

Assumption A2: The agglomeration cost functions are "ordered:" $\Gamma_{m+1}(x) > \Gamma_m(x)$ for any scale x.

Proposition 1 *If A2 is satisfied, then all cities differ in their services prices, π_m.*

Proof. Consider two cities with different Γ_m functions. There are two cases. (i) The cities differ in their χ_m levels. Due to costless labor mobility, the city with higher agglomeration cost must pay higher wages, $w_{H,m}$ and $w_{L,m}$. Firms can only break even, if the high wage city has a low π_m. (ii) The cities share the same χ_m levels. This leads to a contradiction. Due to labor mobility, firms in both cities must pay the same wages. For final goods firms to break even, π_m must be equal in both cities. Thus all factor prices and all factor input ratios are equalized. The cities also have the same $Y_{m,x}$; otherwise the unit costs of services would differ (see Proposition 9). Now consider the identity

$$H_m = H_{m,x} \frac{Y_{m,x}}{H_{m,x}} \frac{H_m}{Y_{m,x}} \tag{20}$$

Note that both cities have the same $X_{m,i}/H_{m,i} = \Psi(\pi_m)$ in all industries, which thus equals $Y_{m,x}/H_m$. Hence, both cities have the same H_m and thus the same size, $H_m + L_m$. This contradicts the assumption that the χ_m are the same. ∎

The intuition is as follows. Consider two cities with identical π_m. All factor price ratios are the same in these cities. Hence, the unit costs of final goods firms are proportional to $w_{H,m}$. Thus, firms must pay the same wages in all cities. Labor mobility then requires that agglomeration costs be the same. If agglomeration cost functions satisfy A2, this means that the cities must have different total employment. Then they also differ in the scale of services and thus in π_m - a contradiction.

3.2 Industry attainment

This section derives the model's implications for educational attainment at the city-industry level, $\lambda_{m,i}$. Since services inputs are not observable, it is necessary to "maximize out" $X_{m,i}$ before taking the model to the data. This yields a reduced form production function that expresses $Y_{m,i}$ as a function of $H_{m,i}$ and $L_{m,i}$. Proposition 2 establishes a key property of this production function: each city is endowed with an *industry neutral* skilled labor augmenting parameter ψ_m. That is, the reduced form production function exhibits skill bias differences across cities that affect all industries symmetrically.

Proposition 2 *Cities are endowed with reduced form production functions of the form*

$$Y_{m,i} = F_i(\psi_m H_{m,i}, L_{m,i}) \tag{21}$$

$$= \left[\mu_i (\psi_m H_{m,i})^\rho + (1 - \mu_i) L_{m,i}^\rho \right]^{1/\rho} \tag{22}$$

where $\psi_m = \psi(\pi_m)$ and $\psi' < 0$.

Proof. Write the production function for final goods as

$$Y_{m,i} = F_i(g[X_{m,i}/H_{m,i}] H_{m,i}, L_{m,i}) \tag{23}$$

Since the optimal $X_{m,i}/H_{m,i} = \Psi(\pi_m)$ is monotone in π_m, we can write

$$g(X_{m,i}/H_{m,i}) = \psi(\pi_m)$$

with $\psi' < 0$. ∎

A similar finding has been documented in the international trade literature. Trefler (1993) finds that countries are endowed with industry neutral factor augmenting technological differences.

Lacking data on industry output, the empirical evaluation of the model relies on the firm's first order conditions to predict industry-city attainment. Proposition 3 shows that these first-order conditions are also characterized by *industry neutral* skill bias differences. That is, the firm's choice of $H_{m,i}/L_{m,i}$ is consistent with a production function that exhibits industry neutral differences in skill bias. However, the skill bias parameters, $\bar{\phi}_m = \phi(\pi_m)$, are different from the ψ_m appearing in the reduced form production function (21).

Proposition 3 *The share of skilled labor employed in city m and industry i is given by*

$$\frac{H_{m,i}}{L_{m,i}} = \left[\frac{\mu_i \bar{\phi}_m}{1 - \mu_i} \frac{w_{L,m}}{w_{H,m}}\right]^{1/(1-\rho)} \tag{24}$$

where $\bar{\phi}_m = \phi(\pi_m) = g(X_{m,i}/H_{m,i})^{\rho-1} G_H(X_{m,i}/H_{m,i})$.

Corollary 4 *Cities with high $\bar{\phi}_m$ employ more educated labor in all industries.*

Corollary 5 *If A1 is satisfied, then $\phi' < 0$ and cities with low services prices (π_m) employ more educated labor in all industries.*

Proof. The final goods firm's first-order conditions imply

$$\begin{aligned}
\frac{w_{H,m}}{w_{L,m}} &= \frac{\mu_i}{1-\mu_i}(G_{m,i}/L_{m,i})^{\rho-1}\ G_H(X_{m,i}/H_{m,i}) \\
&= \frac{\mu_i}{1-\mu_i}(g[X_{m,i}/H_{m,i}]\ H_{m,i}/L_{m,i})^{\rho-1}\ G_H(X_{m,i}/H_{m,i}) \\
&= \frac{\mu_i\ \bar{\phi}_m}{1-\mu_i}\ (H_{m,i}/L_{m,i})^{\rho-1} \tag{25}
\end{aligned}$$

Thus, cities with high $\bar{\phi}_m$ employ more skilled labor in any industry i. Since $X_{m,i}/H_{m,i}$ is a decreasing function of π_m, $\bar{\phi}_m$ and $X_{m,i}/H_{m,i}$ are positively related. A1 then ensures that high $\bar{\phi}_m$ cities also have low π_m. ∎

Proposition 3 establishes one of the main testable prediction of the model. Given estimates of skill premia and of the production function parameters μ_i and $\bar{\phi}_m$, (24) predicts educational attainment for city-industry pairs.

One implication is that, within an industry, attainment differs systematically across cities. Cities with higher $\bar{\phi}_m$ employ more skilled labor in every industry, even though skill premia are equalized across cities. To an econometrician, who does not observe services inputs, cities *appear* to differ in their skill bias parameters $(\bar{\phi}_m)$, even though they are actually endowed with the same production functions. The skill bias differences are *industry neutral* because firms in all industries face the same relative prices (π_m) and use the same production function G.

3.3   Industry specialization

The model implies a second reason why cities with cheap services are more educated: they specialize in skill intensive industries.

Proposition 6 *Cities with low π_m specialize in high μ_i industries. If A1 is satisfied, then cities with high $\bar{\phi}_m$ specialize in industries with high μ_i.*

Proof. The unit cost function of good i is given by

$$c_{m,i} = \left[\mu_i^{1/(1-\rho)} p_{m,G}^{\rho/(\rho-1)} + (1-\mu_i)^{1/(1-\rho)} w_{L,m}^{\rho/(\rho-1)} \right]^{1-1/\rho} \tag{26}$$

where $p_{m,G}(p_{m,x}, w_{H,m})$ is the unit cost of $G(X_{m,i}, H_{m,i})$, which is the same for all industries in a city. Since G is constant returns to scale, we can write:

$$p_{m,G} = w_{H,m} \; c_G(p_{m,x}/w_{H,m})$$

with $c_G' > 0$. The unit cost of good i is then

$$c_{m,i} = w_{L,m} \left[\mu_i^{1/(1-\rho)} \left\{ \frac{w_{H,m}}{w_{L,m}} c_G(\pi_m) \right\}^{\rho/(\rho-1)} + (1-\mu_i)^{1/(1-\rho)} \right]^{1-1/\rho} \tag{27}$$

I show next that unit costs are more sensitive to μ_i when π_m is low. That is

$$\frac{\partial \ln(c_{m,i}/w_{L,m})}{\partial(w_{H,m}/p_{m,x}) \; \partial \mu} < 0 \tag{28}$$

To see (28), define $\hat{\zeta}(w_{H,m}/p_{m,x}) = \left\{ \frac{w_{H,m}}{w_{L,m}} c_G(p_{m,x}/w_{H,m}) \right\}^{\rho/(\rho-1)}$. Note that $\partial \hat{\zeta}/\partial(w_{H,m}/p_{m,x}) > 0$. That is, $c_{m,i}/w_{L,m}$, is low when π_m is low. Write unit cost as

$$\ln(c_{m,i}/w_{L,m}) = \frac{\rho-1}{\rho} \ln\left(\mu_i^\eta \hat{\zeta} + [1-\mu_i]^\eta \right)$$

and take the derivative:

$$\frac{\partial \ln(c_{m,i}/w_{L,m})}{\partial(w_{H,m}/p_{m,x})} = \frac{\rho-1}{\rho} \left[\hat{\zeta} + \left(\frac{1-\mu}{\mu} \right)^\eta \right]^{-1} \hat{\zeta}' < 0$$

The second derivative is negative:

$$\frac{\partial^2 \ln(c_{m,i}/w_{L,m})}{\partial(w_{H,m}/p_{m,x}) \; \partial \mu} < 0$$

Thus, the slope of the unit cost function against μ_i is steeper (negative) when π_m is low. Now consider the ratio of unit costs $c_{m,i}/c_{m,j}$ in two locations. If $\mu_i > \mu_j$, then the low π_m location has a comparative advantage in the high μ_i good. Finally, if A1 holds, low π_m implies high ψ_m and high $\bar{\phi}_m$. ∎

Figure 2 illustrates proposition 6. Consider two cities that differ in their services prices. If wages were equalized across cities, the city with cheaper services would have lower unit cost for all goods. It would have an absolute advantage in all industries and a comparative advantage in high μ_i industries. The intuition is that services prices affect unit costs more in industries that place a large weight on services inputs (μ_i).

Since cities with high $\bar{\phi}_m$ employ more skilled labor within each industry and specialize in skill intensive industries, overall city attainment must be higher as well. Thus, the model can account for the positive relationship between city attainment and $\bar{\phi}_m$ documented in figure 1 in the introduction.

Corollary 7 *If A1 holds, cities with high $\bar{\phi}_m$ are more educated (high H_m/L_m).*
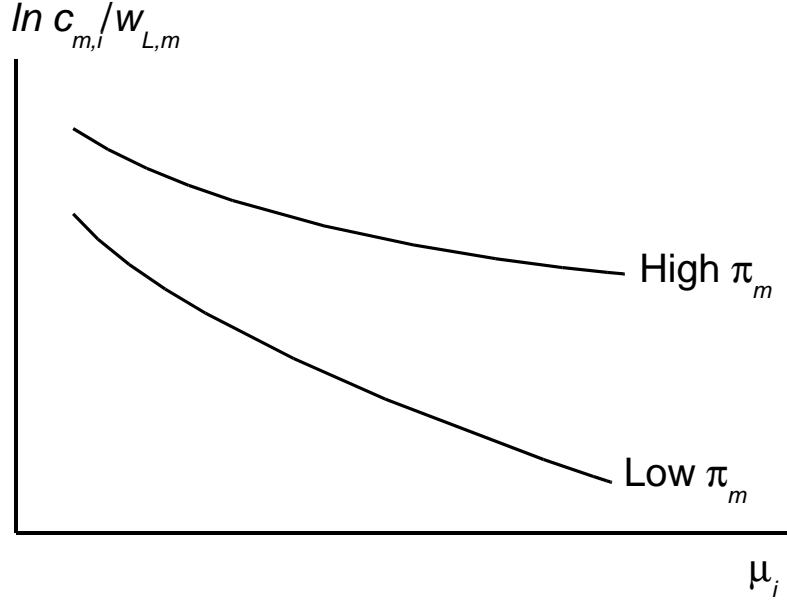
Figure 2: Comparative advantage

Lower services prices increase the scale of the services sector. Hence, the model predicts that cities with low π_m have higher total factor productivity. Lacking data on industry output and services inputs, I cannot test this property directly. However, I can test the prediction that such cities should pay higher wages.

Proposition 8 *If A1 holds, cities with high $\bar{\phi}_m$ pay high wages ($w_{H,m}, w_{L,m}$).*

   Proof. This follows directly from the fact that unit cost is falling in π_m for fixed wages. If A1 holds, then low π_m means high $\bar{\phi}_m$ and unit cost is falling in $\bar{\phi}_m$ as well. As a result, cities with high $\bar{\phi}_m$ have absolute advantage in all goods. To remain competitive, cities with lower $\bar{\phi}_m$ then must pay lower wages (or be shut down).   ∎

3.4   Size of the business service sector

One set of predictions that could distinguish the model proposed here from alternative theories relates to the business services sector. Lacking data on the price of services, I derive how the size of the services sector differs between cities with different attainment (and thus π_m).

Proposition 9 *Cities with low services prices (π_m) have larger business services output $(Y_{m,x})$.*

   Proof. From (27) the unit cost of X_m may be written as

$$p_{m,x} = c_x \left(\pi_m, w_{H,m}/w_{L,m} \right) \ w_{L,m} \ / \ F_x \left(G_{m,x}, L_{m,x} \right)^{\varepsilon} \tag{29}$$

with $\partial c_x / \partial \pi_m > 0$. c_x is the unit cost function for the constant returns to scale production function F_x. Thus

$$\frac{p_{m,x}}{w_{L,m} \ c_x} = \frac{\pi_m}{c_x} \frac{w_{H,m}}{w_{L,m}} = F_x \left(G_{m,x}, L_{m,x} \right)^{-\varepsilon} \tag{30}$$

10

Since the elasticity of unit cost with respect to any input price (here: π_m) is less than one, the ratio on the left-hand-side of (30) is decreasing in π_m for given skill premium. Thus, the model predicts that cities with low π_m have large $Y_{m,x}$. ■

This result is intuitive. In some cities, services are cheap due to economies of scale. This requires large services output. Unfortunately, services output is not observable in my data. I therefore, derive implications for services *employment*, which is observable. This requires an additional assumption.

Assumption A3: $\partial \ln Y_{m,x} / \partial \ln X_{m,x} < 1$.

A3 states that services are produced subject to diminishing returns to X. This assumption is necessary for the model to make sense. Were it violated, it would be possible to produce larger amounts of services using less labor of both skills. If A3 is satisfied, the next proposition shows that cities with cheap services either employ large amounts or large fractions of skilled labor in the production of services.

Proposition 10 *If A1 and A3 hold, then cities with low π_m have either high $H_{m,x}/H_m$ or high $H_{m,x}$.*

Proof. Write the fraction of skilled labor employed in producing services as

$$\frac{H_{m,x}}{H_m} = \frac{H_{m,x}}{Y_{m,x}}\frac{Y_{m,x}}{H_m} = \frac{H_{m,x}}{Y_{m,x}}\frac{X_{m,x}}{H_{m,x}} \tag{31}$$

The last equality holds because $X_{m,i}/H_{m,i}$ is the same in all industries and $\sum X_{m,i} = Y_{m,x}$. Thus

$$\frac{H_{m,x}}{H_m} = \frac{X_{m,x}}{Y_{m,x}} \tag{32}$$

This can be satisfied in two ways. If cities with low π_m have low $H_{m,x}$, then such cities also have low $L_{m,x}$ because $H_{m,x}/L_{m,x}$ is decreasing in π_m due to A1. Due to A3, such cities have high $X_{m,x}/Y_{m,x}$ and thus high $H_{m,x}/H_m$. Alternatively, $H_{m,x}$ may be high in cities with low π_m. The model then makes no predictions about $H_{m,x}/H_m$. ■

Note that the case where cities with cheap services have low $H_{m,x}$ is empirically implausible. It implies that such cities also have low H_m and thus L_m. That is, cities with larger services sectors would have smaller (in absolute terms) employment in all other industries.

3.5   City size

Empirically, the relationship between city population and educational attainment is positive, but weak (see Glaeser 1999 and the evidence presented in section 4.7). This seems to pose a challenge for models based on agglomeration economies. However, in my model, the correct measure of city "size" is not its total population, but the scale of the business services sector. In contrast to Eaton & Eckstein (1997) and Glaeser (1999), my model does not imply that cities with larger population sizes are more educated. Proposition 11 identifies the reason for this: agglomeration costs function (Γ_m) differ across cities.

Proposition 11 *If all cities share the same Γ_m, then cities with low services prices employ more labor.*

Proof. If all cities share the same Γ_m, then larger cities pay higher wages. They must have lower unit costs, $c_{m,i}/w_{H,m}$. This requires low π_m. ■

4   Empirical Evaluation of the Model

The model's empirical implications may be summarized as follows: If A1 through A3 hold, then cities with cheap services are characterized by high values of $\bar{\phi}_m$, high city attainment H_m/L_m, high attainment in each industry $H_{m,i}/L_{m,i}$, high wages, and by large employment shares in skill intensive industries. In addition, such cities have large absolute ($H_{m,x}$) or relative ($H_{m,x}/H_m$) skilled employment in business services. This section shows that all of the model's predictions have strong support in data on U.S. cities.

4.1   Data

The data are drawn from the 1980 to 2000 waves of the Decennial Census 5% Public Use Micro Samples. Individuals are dropped if they satisfy at least one of the following conditions: reside in group quarters, in school, zero hours worked, younger than age 17 or older than age 75. A person with at least 16 years of schooling is classified as holding a college degree. I refer to such persons as "skilled." Individual wage rates are calculated as the ratio of labor income to hours worked per year. Wage observations below 10% of the median and above 100 times the median are deleted as likely measurement errors.

   The geographic units considered are metropolitan areas (MAs). Most of the findings reported in this paper are based on data for the year 2000. Results for earlier years are similar, except where explicitly noted. Appendix A provides more detail on the data and variable construction.

4.1.1   Measurement

Labor input.   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

$$\ln h\left(e_j, x_j, z_j\right) = \beta_0 + \beta_1 e_j + \beta_2 x_j + \beta_3 x_j^2 + \beta_4 z_j + \varepsilon_j \tag{33}$$

where e denotes years of schooling, x is experience, and z denotes other demographic characteristics (race and sex). Labor efficiency (h) 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

$$H_{m,i} = \sum_j h\left(e_j, x_j, z_j\right) \ l_j \tag{34}$$

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. City-industry cells with fewer than 20 observations are dropped.

   Skill premia are calculated as the ratio of skilled to unskilled wage bills divided by H/L.

Production functions.   The production function parameters to be estimated are the μ_i for all industries, the $\bar{\phi}_m$ for all cities, and ρ.

   The parameter ρ governs the substitutability of G and L. Direct evidence on its value does not exist. Empirical estimates of the substitution elasticity between skilled and unskilled labor are typically based on aggregate rather than industry data (see Ciccone and Peri 2004). Moreover, how strongly $H_{m,i}/L_{m,i}$ responds to a change in $w_{H,m}/w_{L,m}$ depends on the accompanying change in π_m. Lacking better evidence, I set ρ such that the elasticity of substitution between G and L

Table 1: Business services

Code	Industry
716	Banking and credit
726	Security and commodity brokerage and invest companies
736	Insurance
746	Real estate
756	Real estate-insurance-law offices
806	Advertising
807	Accounting, auditing, and bookkeeping services
808	Misc business services
879	Legal services
898	Engineering and architectural services

Notes: The table shows the industries classified as business services.

is $(1-\rho)^{-1} = 1.6$. This is in the middle of the empirical range cited by Ciccone and Peri (2004). The sensitivity analysis explores alternative values of ρ.

The values of μ_i and $\bar{\phi}_m$ are estimated from the first-order conditions of the final goods firm (24). The model implies

$$\ln\left(H_{m,i}/L_{m,i}\right) = \frac{\ln\left(\mu_i/\left(1-\mu_i\right)\right) + \ln\left(\bar{\phi}_m\right) - \ln\left(w_{H,m}/w_{L,m}\right)}{1-\rho} \tag{35}$$

This translates into a regression equation

$$\ln\left(H_{m,i}/L_{m,i}\right) = \beta_0 + D_i + \bar{D}_m + \beta_w \ln\left(w_{H,m}/w_{L,m}\right) + \varepsilon_{m,i} \tag{36}$$

which I estimate via OLS. The skill bias parameters may then be recovered from the industry dummies (D_i)

$$\mu_i = \frac{e^{(1-\rho)(\beta_0+D_i)}}{1 + e^{(1-\rho)(\beta_0+D_i)}} \tag{37}$$

and from the state dummies $\left(\bar{D}_m\right)$:

$$\bar{\phi}_m = e^{\bar{D}_m(1-\rho)} \tag{38}$$

Below, estimates of the production function parameters are used to predict city-industry attainment according to the firm's first-order condition (24). In these calculations, an alternative measure of $\bar{\phi}_m$ is used. A key result of the paper is that employment in business services predicts a large share of the variation in educational attainment across cities. To drive home this point, it is desirable to estimate the skill bias parameters $\bar{\phi}_m$ from data on business services only. This is done by solving the business services firm's first order condition (24) for $\bar{\phi}_m$, given data for skill premia, $H_{m,x}/L_{m,x}$, and the values of μ_i taken from (37). This alternative measure of skill bias is called ϕ_m and is used for most of the paper's results. Note that $\ln\left(\phi_m\right)$ is a linear transformation of $\ln\left(H_{m,x}/L_{m,x}\right)$.

Industries. Each worker is assigned to one of 150 industries according to the IPUMS variable IND1950. Table 1 lists the industries classified as business services. Some of these industries serve consumers as well as businesses. Unfortunately, it is not possible to measure what fraction of output a given industry supplies to businesses. For some industries, the classification as business services is therefore to some extent arbitrary.

Table 2: City characteristics

	1980	1990	2000
N	258	271	297
H_m/L_m [%]	24.2	31.6	38.5
	(8.7)	(12.5)	(16.5)
$H_{m,x}/H_m$ [%]	13.8	15.5	18.2
	(4.4)	(4.9)	(6.4)
$L_{m,x}/L_m$ [%]	8.3	9.9	11.2
	(2.5)	(2.8)	(3.1)
Frac.services [%]	10.1	12.2	14.0
	(3.0)	(3.6)	(4.3)
$w_{H,m}/w_{L,m}$	1.52	1.59	1.59
	(0.10)	(0.08)	(0.08)

Notes: The table shows means and standard deviations (in parentheses) for MA characteristics. Frac. services is the fraction of workers employed in business services. N is the number of cities.

Table 3: Attainment and skill bias

Year	1980	1990	2000
β	1.08	1.27	1.27
s.e.	0.06	0.06	0.04
R^2	0.54	0.62	0.78
N	258	271	297

Notes: The table shows the results of regressing $\ln(H_m/L_m)$ on $\ln(\phi_m)$. β is the regression coefficient and "s.e." denotes its standard error.

4.1.2 Summary statistics

Table 2 shows means and standard deviations of MA characteristics for each year. College attainment (H_m/L_m) rises by nearly 60% over the 20 year sample period. The fraction of skilled and unskilled labor employed in business services increases by around one-third. Data for individual cities are shown in the Appendix.

4.2 Skill bias and education

The following sections confront the empirical implications derived in section 3 with the data. Corollary 7 is the main qualitative prediction of the model: Cities with high ϕ_m, reflecting low services prices, are highly educated. Figure 1, displayed in the introduction, shows that this prediction is strongly supported in the data.

Variation in ϕ_m accounts for a large part of the cross-city variation in attainment. A regression of $\ln(H_m/L_m)$ on $\ln(\phi_m)$ yields an R^2 of 0.78 for the year 2000. Table 3 shows the regression results for the years 1980 to 2000. The correlation between attainment and ϕ_m is strong in all years, but substantially lower in earlier years.

It is worth reiterating that ϕ_m is estimated solely from data on business services employment. It does not use any information about other city characteristics that could plausibly account for education gaps, including city size (suggested by agglomeration theories), industry specialization, skill premia, or geographic location. The strong correlation between ϕ_m an city education is consistent with the hypothesis of the paper: business services affect the demand for skilled labor in a city.

The model offers two reasons why cities with high ϕ_m are educated: (i) they employ more skilled labor in all industries and (ii) they specialize in skill-intensive industries. Next, I explore the evidence for both reasons.

4.3 City-industry attainment

Corollary 4 predicts that cities with higher ϕ_m should employ more skilled labor in each industry. Figure 3 examines this prediction informally. For the eight largest industries, the figure plots $\ln(H_{m,i}/L_{m,i})$ against $\ln(\phi_m)$. Least squares regression lines (solid) and 45-degree lines (dashed) are also shown. Consistent with the model prediction, a strong correlation is visible in all cases.[2]

This patterns holds for the full set of industries. Regressing $\ln(H_{m,i}/L_{m,i})$ against $\ln(\phi_m)$ for each of the 49 largest industries yields positive slope coefficients in every case. The mean slope is 0.52 with a standard deviation 0.21. The mean R^2 across industries is 0.33. That is, ϕ_m accounts for one-third of the variation in attainment within industries across cities.

Quantitative evaluation. The model's Proposition 3 makes quantitative predictions for attainment in city-industry pairs. These can be used to measure the fraction of attainment variation accounted for by the model.

Figure 4 plots the model's predicted attainment against observed attainment for the year 2000. Predicted attainment is calculated from the firm's first-order conditions (24) together with estimates of μ_i and ϕ_m. The U.S. skill premium is used for $w_{H,m}/w_{L,m}$. An OLS regression yields an R^2 of 0.77. Thus, the model accounts for 77% of the observed variation in attainment across city-industry pairs.

Attainment variation within industries. Part of the attainment variation shown in figure 4 is due to technological differences across cities (μ_i) about which the model has nothing to say. It is therefore more appropriate to evaluate the model by its ability to account for attainment gaps across cities *within industries*.

According to the model, attainment net of industry effects is given by

$$\lambda_{m,i}^{net} = \ln(H_{m,i}/L_{m,i}) - \frac{\ln(\mu_i/(1-\mu_i))}{(1-\rho)} \tag{39}$$

Equation (24) predicts that $\lambda_{m,i}^{net}$ should equal

$$\lambda_{m,i}^{pred} = \ln(w_{L,m}/w_{H,m}) + \frac{\ln(\phi_m)}{1-\rho} \tag{40}$$

Note that predicted attainment, net of industry effects, varies solely due to ϕ_m. Figure 5 evaluates this prediction by plotting $\lambda_{m,i}^{pred}$ against $\lambda_{m,i}^{net}$. A fitted OLS regression line,

$$\lambda_{m,i}^{pred} = \beta_0 + \beta\lambda_{m,i} + \varepsilon_{m,i} \tag{41}$$

yields a slope coefficient of 0.5, compared with a model prediction of 1. In this sense, the model accounts for 50% of the variation in attainment, within industries, between high and low ϕ cities. The regression $R^2 = 0.29$. Thus, the model accounts for 29% of the variation in attainment across all city-industry pairs, net of industry effects.

In evaluating the model's performance it is worth keeping in mind that predicted attainment is calculated only from data on business services employment (ϕ_m). In addition, part of the variation in observed attainment is due to measurement error as many of the industry-city cells contain only small numbers of observations.

---

[2]Lewis (2004) and Glaeser and Berry (2005) show that cities absorb inflows of labor mostly by changing attainment within industries, not by changing industry mix. Beaudry et al. (2005) and Lewis (2005) argue that this happens through local technology adoption.
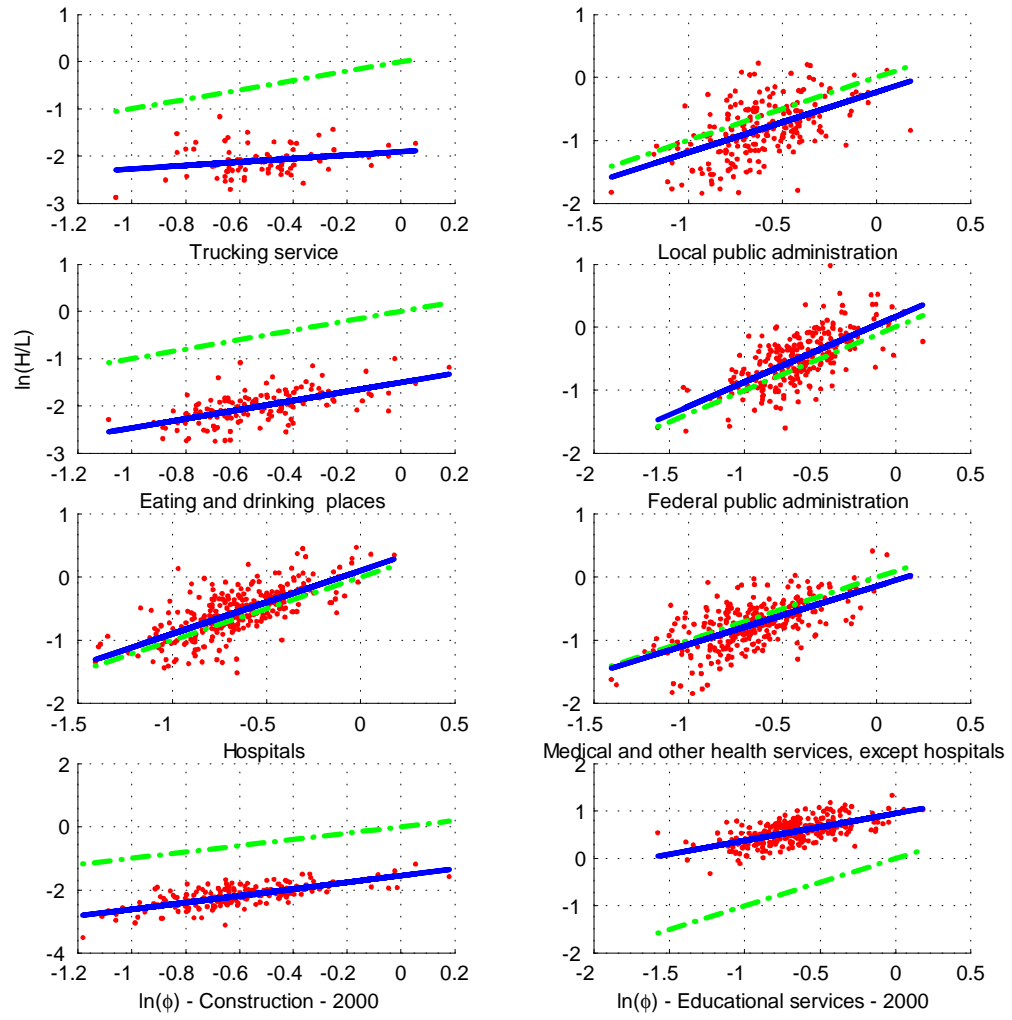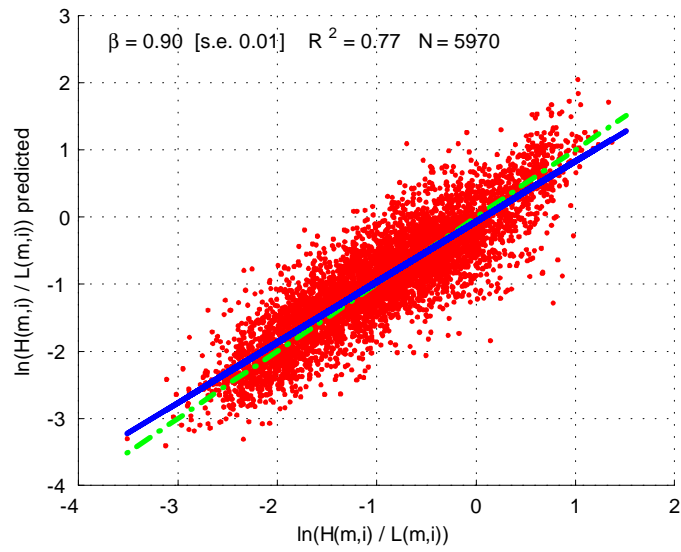
Figure 3: Attainment and skill bias

16

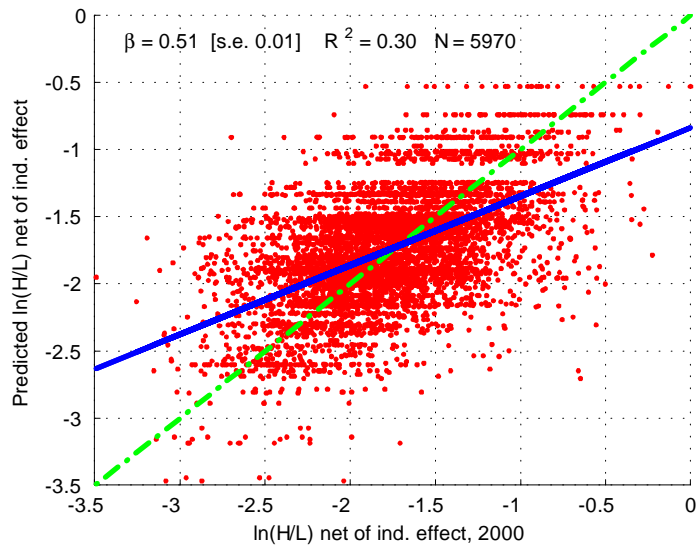Figure 4: Predicted and observed attainment

Figure 5: Attainment net of industry effects

4.4 Industry Specialization

Proposition 6 points to a second reason why cities with high ϕ are highly educated: they specialize in skill-intensive industries. To evaluate this prediction, I decompose city attainment gaps into the contributions of industry specialization and within-industry attainment differences.

The starting point is the identity

$$\frac{H_m}{L_m} = \frac{\sum_i H_{m,i}}{\sum_i L_{m,i}} = \sum_i \frac{H_{m,i}}{L_{m,i}} \, \omega_{m,i} \tag{42}$$

where $\omega_{m,i} = L_{m,i}/L_m$ is the employment share of industry i. The identity suggests a natural way of decomposing attainment gaps into the contributions of industry specialization ($\omega_{m,i}$) and within industry attainment differences ($H_{m,i}/L_{m,i}$). Define two versions of counter-factual city attainment:

$$\eta_m^{IS} = \sum_i \frac{H_{US,i}}{L_{US,i}} \, \omega_{m,i} \tag{43}$$

$$\eta_m^{\phi} = \sum_i \frac{H_{m,i}}{L_{m,i}} \, \omega_{US,i} \tag{44}$$

where $\omega_{US,i} \equiv L_{US,i}/L_{US}$ is the economy wide employment share of industry i. $H_{US,i} \equiv \sum_m H_{m,i}$ and $L_{US,i} \equiv \sum_m L_{m,i}$ denote economy wide total employment in industry i. η_m^{IS} eliminates within-industry differences in attainment and thus measures the importance of industry specialization. η_m^{ϕ} eliminates differences in industry specialization and thus measures the importance of within-industry attainment differences.

One complication encountered in the computation of η_m^{IS} and η_m^{US} is that a given city has (near) zero employment in many industries.[3] Conversely, many industries are present only in a few cities. In computing η_m^{IS} and η_m^{US} I therefore drop cities that have data for fewer than 30 industries and industries that have data for fewer than 12 cities. The resulting sample consists of 142 cities and 75 industries. 69% of the remaining industry-city pairs have data.

Figure 6 plots predicted against observed state attainment. The slope is 0.20, leading me to conclude that specialization accounts for 20% of attainment differences.[4]

In a similar fashion, the importance of within industry differences in attainment can be quantified. Recall that η_m^{ϕ} applies the same employment weights to all cities. Its variation is solely due to within-industry attainment differences across cities.

Figure 7 plots counter-factual city attainment against observed attainment for the year 2000. The slope of an OLS regression line is 0.85 (*s.e.* 0.02). In this sense, within-industry attainment differences account for a large majority of city attainment gaps.

4.5 Business Services

This section examines the empirical relationship between business services and city attainment. This evidence is useful for distinguishing the proposed theory from possible alternatives in which business services play no special role. Lacking data on the quantity ($X_{m,i}$) and price of services (π_m), I evaluate the model's predictions for services employment.

Proposition 10 predicts a positive correlation between ϕ_m and either the level ($H_{m,x}$) or the share ($H_{m,x}/H_m$) of skilled employment in services. Both correlations are found in the data (see

---

[3] Rosenthal and Strange (2002) show that this is true even at the state level.

[4] The model's demand side is not sufficiently developed to compare the model's predicted industry composition with the data.
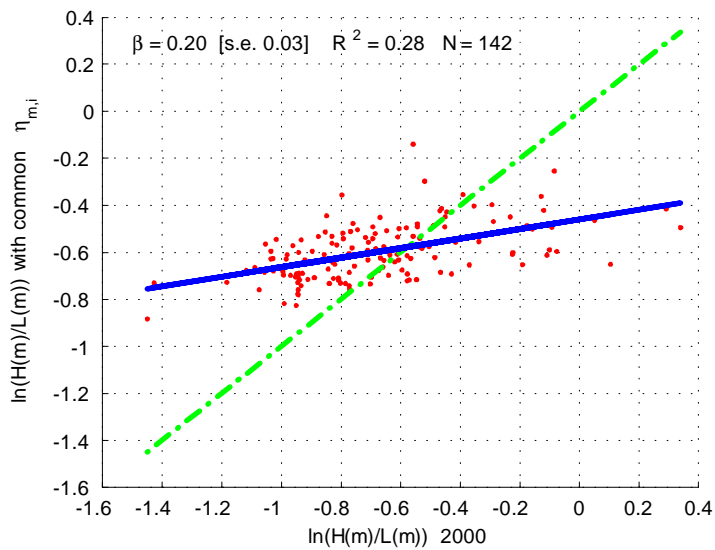
18

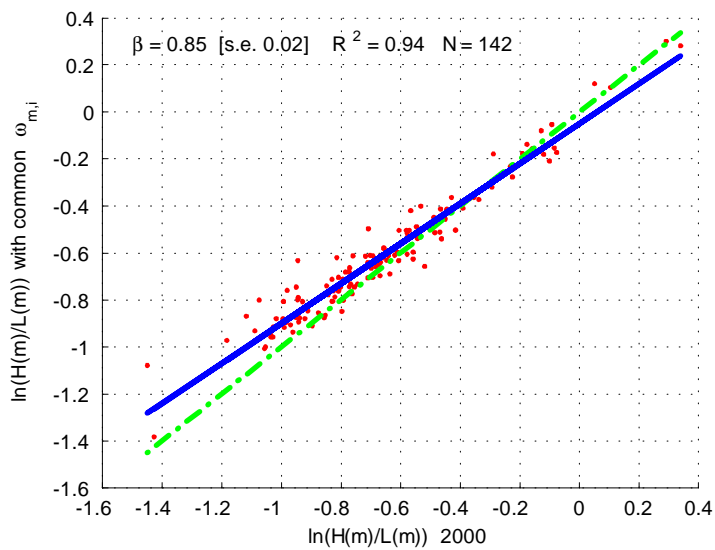Figure 6: Importance of industry specialization

Figure 7: Importance of skill bias differences

19

figures 8 and 9). The finding that highly educated cities employ a larger *share* of labor in business services confirms that the positive association between ϕ_m and $H_{m,x}$ is not simply due to city size.
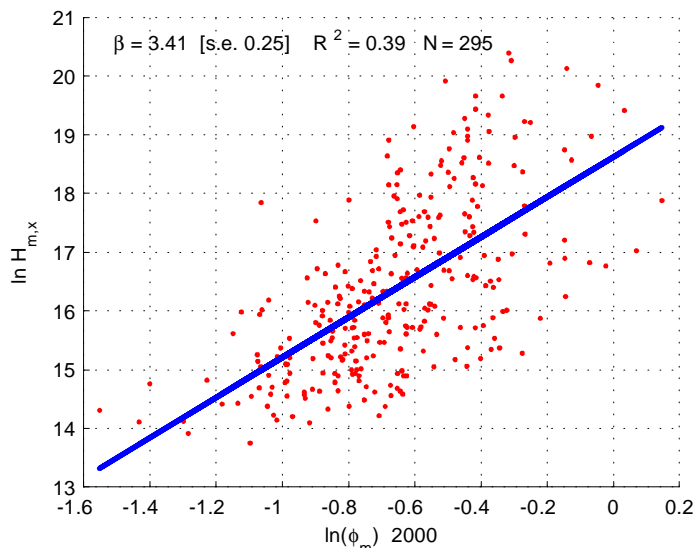

Figure 8: Skill bias and business services

4.6 Wages

According to Proposition 8, cities with higher ϕ_m pay higher nominal wages. To investigate this proposition, I calculate a city's mean wage as

$$\bar{w}_m = \sum_j w_{m,j} \omega_{US,j} \tag{45}$$

where j indexes age / education / sex groups and $\omega_{US,j}$ is the U.S. population weight. $\bar{w}_m$ removes differences in demographic composition from the mean wages of cities.

Consistent with the model's prediction, I find a positive, albeit weak, correlation between ϕ_m and $\bar{w}_m$. Figure 10 shows a scatter plot. An OLS regression line yields an R^2 of 0.16. This is consistent with Rauch's (1993) finding of a positive relationship between education and wages across cities.

I also find a significant and positive correlation between a city's mean wage and the size of the business services sector, measured by $\ln(H_{m,x})$ or $\ln(H_{m,x}/H_m)$. This is consistent with the notion that business services account for productivity and thus for wage differences across cities.

4.7 City size

It is well known that larger cities are more educated (e.g., Glaeser 1999). This motivates agglomeration theories in which human capital externalities facilitate learning in cities (Eaton & Eckstein 1997; Glaeser 1999).

Table 4 investigates the relationship between city size and education in my data. The correlation is positive in all years, but it is not strong. Regressing $\ln(H_m/L_m)$ on the logarithm of city population yields statistically significant point estimates. However, the R^2 are only on the order of 10%. This is consistent with the predictions of my model, in which the correct measure of
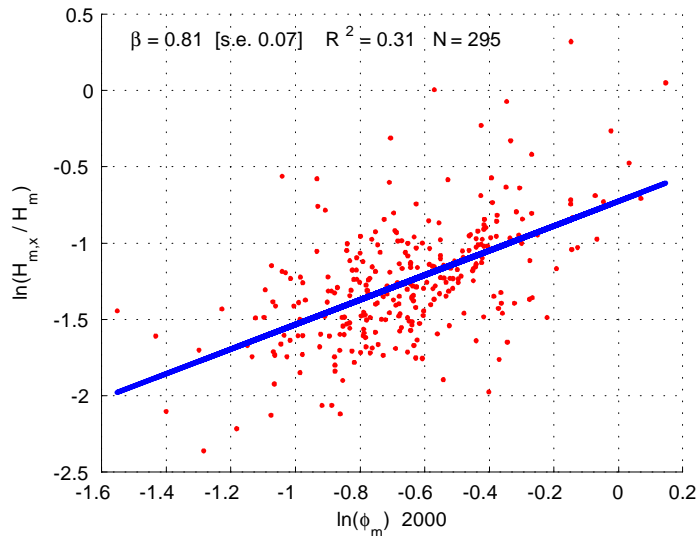
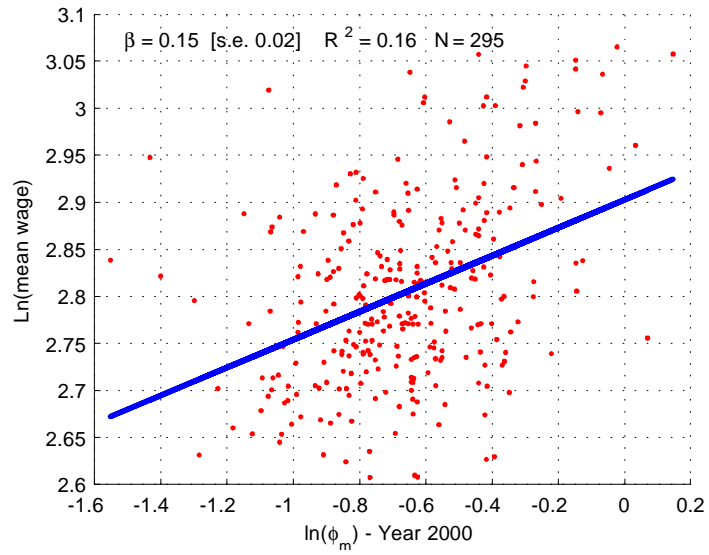Figure 9: Skill bias and size of business services sector.

Figure 10: Wages and skill bias

Table 4: Attainment and city population

	1980	1990	2000
β	0.115	0.136	0.155
s.e.	0.022	0.023	0.024
R^2	0.10	0.12	0.12
N	240	247	289

Notes: The table shows the results of regressing $\ln\left(H_m/L_m\right)$ on the logarithm of city population. β is the regression coefficient and "s.e." denotes its standard error.

Table 5: Attainment and skill premia at the city level

Year	1980	1990	2000			
β	0.032	-0.008	-0.004			
s.e.	0.011	0.009	0.007			
R^2	0.03	0.00	0.00			
N	271	273	297			
$w_{H,m}/w_{L,m}$	1.52	1.59	1.59			
Std	0.10	0.08	0.08			

Notes: The table shows the results of regressing $\ln\left(w_{H,m}/w_{L,m}\right)$ on $\ln\left(H_m/L_m\right)$. β is the regression coefficient and "s.e." denotes its standard error. The last two columns show the mean and standard deviation of skill premia across cities.

agglomeration is the scale of the business services sector, not city population. As shown in section 4.5, the correlation between the scale of services and city education is substantially stronger than the correlation between population size and education.

4.8 Skill Premia

In the model, labor mobility equalizes skill premia across locations and industries. An alternative view holds that labor is relatively immobile and that the supply of labor is an important determinant of city attainment. [reference] If skills are imperfect substitutes, such theories imply that highly educated cities should have low skill premia. This section examines whether skill premia are correlated with attainment at the city and city-industry level.

Table 5 shows the results for city level data. I regress the logarithm of city skill premia on city attainment for the years 1980 to 2000. The slope coefficient varies in sign and is significant only in 1980. The R^2 statistics are at most 2%.

Table 6 shows that similar results are obtained for city-industry pairs. I regress the logarithm of the skill premium on industry dummies and attainment.

$$\ln\left(w_{H,m,i}/w_{L,m,i}\right) = D_i + \beta \ln\left(H_{m,i}/L_{m,i}\right) + \varepsilon_{m,i} \tag{46}$$

The slope coefficients are small and change signs across years. A 10% change in attainment is associated with a change in the skill premium of roughly 0.1%. When industry dummies are omitted, the slope coefficients are always positive and slightly larger. However, the regression R^2 drops to less than 3%. This is consistent with Peri's (1998) findings of a small, positive correlation between skill premia and schooling across city-industry cells.

I conclude that skill premia are, to first approximation, uncorrelated with educational attainment. This is consistent with the assumption of perfect labor mobility underlying the model.

22

Table 6: Attainment and skill premia at the city-industry level

Year	Dummies	β	s.e.	R^2	N
1980	yes	0.005	0.005	0.464	3810
1980	no	0.028	0.003	0.022	3810
1990	yes	-0.010	0.004	0.406	4846
1990	no	0.033	0.003	0.031	4846
2000	yes	-0.013	0.004	0.404	5970
2000	no	0.025	0.002	0.016	5970

Notes: The table shows the results of estimating (46) via OLS. "s.e." denotes the standard error of β.

5   Conclusion

This paper explores why educational attainment differs across U.S. metropolitan areas. A measure of business services employment (ϕ_m) is identified, which predicts a large share of the observed cross-city variation in attainment. This motivates a theory in which the scale of a city's business services sector affects its demand for skilled labor across all industries. The model makes a number of empirically testable prediction, all of which find strong support in the data.

Appendix

A Data Sources and Variable Construction

Table A1 shows summary statistics for the ten most educated and least educated cities. Some of the highly educated cities, such as Ann Arbor or Madison, host large universities. These might increase the supply of college educated labor. I do not drop such cities for two reasons. (i) The model assumes perfect labor mobility. A city's attainment should therefore not be correlated with its production of college graduates. (ii) It is conservative to retain cities with idiosyncratic variation in attainment. They pose a challenge for the model and should reduce its explanatory power.

Table A1: Summary statistics for the most and least educated cities.

Name	Code	H_m/L_m	$ln(\phi)$	Mean wage	$\frac{w_{H,m}}{w_{L,m}}$	Population	Density	N
Most educated								
San Francisco-Oakland-Val	7360	1.16	-0.28	18.52	1.60	1731.2	1704.7	41.6
Boston, MA	1120	0.92	-0.42	17.73	1.58	3406.8	1685.1	77.3
San Jose, CA	7400	0.88	-0.86	19.81	1.69	1682.6	1303.6	39.2
Ann Arbor, MI	440	0.81	-0.14	17.69	1.49	578.7	285.3	8.8
Washington, DC/MD/VA	8840	0.80	-0.31	18.50	1.60	4923.2	756.3	103.1
Middlesex-Somerset-Hunter	5604	0.78	-0.53	19.38	1.60	1169.6	1120.0	29.6
Raleigh-Durham, NC	6640	0.78	-0.70	16.01	1.66	1187.9	340.5	29.7
Madison, WI	4720	0.74	-0.34	16.20	1.48	426.5	354.9	7.9
Fort Collins-Loveland, CO	2670	0.70	0.07	14.75	1.53	251.5	96.7	4.9
Seattle-Everett, WA	7600	0.70	-0.37	17.11	1.48	2414.6	545.9	52.2
Least educated								
Redding, CA	6690	0.20	-0.05	14.93	1.59	163.3	43.1	3.8
Riverside-San Bernadino,	6780	0.18	-1.21	16.44	1.55	3254.8	119.4	57.6
Modesto, CA	5170	0.16	-0.69	16.89	1.55	447.0	299.2	9.3
Stockton, CA	8120	0.16	-1.19	17.13	1.58	563.6	402.8	11.3
Bakersfield, CA	680	0.14	-1.08	16.29	1.71	661.6	81.3	11.9
Yakima, WA	9260	0.14	-1.02	15.31	1.62	222.6	51.8	4.5
Yuba City, CA	9340	0.14	-0.73	15.43	1.60	139.1	112.8	3.2
Ocala, FL	5790	0.14	-0.55	13.23	1.77	258.9	164.0	5.0
Visalia-Tulare-Portervill	8780	0.13	-1.00	15.44	1.67	368.0	76.3	9.1
Merced, CA	4940	0.11	-1.08	15.65	1.62	210.6	109.2	4.1

Notes: The table shows the summary statistics for the most and the least educated cities for the year 2000. Population is taken from U.S. Census Bureau estimates of total MA population (in thousands). Density is the ratio of population to land area (in thousands of person per square mile). N is the number of observations in the PUMS sample.

References

[1] Beaudry, Paul; Mark Doms; Ethan Lewis (2006). "Endogenous skill bias in technology adoption: city-level evidence from the IT revolution." NBER Working Paper #12521.

[2] Berry, Christopher R.; Edward Glaeser (2005). "The divergence of human capital levels across cities." Mimeo. Harvard University.

[3] 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.

[4] Doms, Mark; Ethan Lewis (2005). "Information technology diffusion, human capital, and spillovers: PC diffusion in the 1990s and early 2000s." Mimeo. Federal Reserve Bank of San Francisco.

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

[6] Eaton, Jonathan; Zvi Eckstein (1997). "Cities and growth: theory and evidence from France and Japan." *Regional Science and Urban Economics* 27(4-5): 443-74.

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

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

[9] 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.

[10] Lewis, Ethan (2004). "Local open economies within the U.S.: how do industries respond to immigration?" Working paper 04-1. Federal Reserve Bank of Philadelphia.

[11] Lewis, Ethan (2005). "Immigration, skill mix, and the choice of technique." Working paper. Federal Reserve Bank of Philadelphia.

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

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

[14] Rosenthal, Stuart S.; William C. Strange (2002). "Evidence on the nature and sources of agglomeration economies." Mimeo. Syracuse University.

[15] Trefler, Daniel (1993). "International factor price differences: Leontief was right!" *Journal of Political Economy* 101(6): 961-87.