

Identifying Human Capital Externalities: Theory with an Application to US Cities

Antonio Ciccone*

and

Giovanni Peri

March 2002

Abstract: Identification of the strength of human capital externalities at the aggregate level is still not fully understood. The existing method may yield positive or negative externalities even if wages reflect marginal social products. We propose an approach that yields positive average human capital externalities if and only if the marginal social product of workers with above-average human capital exceeds their wage. As an application, we estimate the strength of average-schooling externalities in US cities between 1970 and 1990.

Key Words: Marginal social product of human capital, Wages, Human capital externalities, Imperfect substitution, Perfect substitution, Cities.

JEL Codes: O0, O4, R0, J3.

*Universitat Pompeu Fabra, CEPR, and IZA, and University of California, Davis respectively. We thank David Card for help with the data, Douglas Almond for excellent research assistance, and Daron Acemoglu, Orazio Attanasio, Joshua Angrist, Richard Blundell, Ken Chay, Adriana Kugler, Costas Meghir, Jonathan Temple, and Gianluca Violante for comments. Daron Acemoglu and Joshua Angrist kindly made their data and programs available to us. The theoretical results on identification in this paper were in part contained in two previous working papers, "Human Capital and Externalities in Cities," (CEPR Discussion Paper No. 2599, 2000) by Ciccone and Peri and "Capital, Wages, and Growth: Theory and Evidence" by Ciccone, Peri, and Almond (CEPR Discussion Paper 2199, 1999). Ciccone thanks the Commission on Research at UC Berkeley (1997-1998), the Spanish Ministry of Education (1997-1999), and CREI for financial support for this project. Corresponding author: Antonio.Ciccone@Econ.Upf.Es.

1 Introduction

Human capital externalities at the aggregate level play a central role in applied economic theory as well as economic policy analysis. In applied theory, human capital externalities are invoked to capture key features of the data (e.g. Lucas (1988), Azariadis and Drazen (1990), Black and Henderson (1999)). In policy analysis, the strength of human capital externalities is one of the main determinants of the optimal subsidy to human capital (e.g. Gemmell (1997), Heckman and Klenow (1998), Heckman (2000)). Assessing the strength of human capital externalities at the aggregate level is therefore important for economic theory as well as policy, and empirical research has responded with a variety of different approaches and estimates (e.g. Rauch (1993), Conley, Flier, and Tsang (1999), Acemoglu and Angrist (2000), Moretti (2000), Rudd (2000)). The theoretical identification problem is still not fully understood however. The difficulty is simple to explain. Empirical work finds that workers with different levels of education are imperfect substitutes in production (e.g. Katz and Murphy (1992), Johnson (1997), Topel (1997), Autor, Katz, and Krueger (1998), Card and Lemieux (2000)). An increase in the aggregate supply of highly educated workers will therefore tend to increase wages of workers with low levels of education and decrease wages of workers with high levels of education, even if wages of highly educated workers reflect their marginal social product (and there is no need for corrective policies). Can we avoid mistaking these standard supply effects with (positive or negative) human capital externalities at the aggregate level?

So far the answer to this question is unclear as existing work on the estimation of the strength of human capital externalities at the aggregate level assumes that workers with different human capital are perfect substitutes in production. Perfect substitutability simplifies the theoretical identification problem because it implies that the aggregate supply of human capital does not affect individual wages if there are no externalities. All effects of the supply of human capital on individual wages can therefore be interpreted as externalities. This yields, for example, that average-schooling externalities at the local geographic level can be estimated by simply including average schooling of the local workforce in a standard Mincerian wage regression (e.g. Rauch (1993), Acemoglu and Angrist (2000), Rudd (2000)). It can be shown however that if workers with different levels of education are imperfect substitutes in production then this (Mincerian) approach to human capital externalities at the aggregate level may yield positive *or* negative externalities even if wages reflect marginal social products.

We therefore propose an approach to the identification of human capital externalities at the aggregate level that is theoretically valid whether workers with different human capital are perfect or imperfect substitutes in production. The approach can be applied at the city-level, the region-level, or (with some modifications that will be explained later) the country-level. The theoretical basis of the approach at the local geographic level is that if wages $w(Z)$ of workers with skills Z reflect marginal social products then changes in the average level of human capital of the local workforce have no first-order effect on average wages $w = \sum_z w(Z)l(Z)$ when workforce skill-composition weights $l(Z)$ are held constant.¹ If wages of high-skilled workers are below their marginal social product however then there will be a positive first-order effect of average human capital at the aggregate level on average wages even when workforce skill-composition is held constant.

To get some intuition of how this theoretical argument applies to average-schooling externalities at the city-level, consider a city experiencing an inflow of highly educated workers resulting in a small increase in average schooling and hence productivity (we will deal with large increases later). For simplicity, assume that output is produced without physical capital and land. Suppose also that workers with different levels of education are imperfect substitutes in production and that there are no city-level average-schooling externalities (wages reflect marginal social products). In this case, incoming workers, through their effect on the supply of different levels of education in the city, raise the wage for some levels of education and lower it for others. But because they are paid their marginal social product, incoming workers will not affect total wage income of the group of workers who already were in the city before the human capital inflow. Assuming for simplicity that productivity does not depend on the aggregate scale of production (we will account for possible scale effects later), this implies that average wages using the city's workforce skill-composition before the human capital inflow will be the same before and after the increase in average schooling. Now suppose instead that there are positive city-level average-schooling externalities. Wages of incoming, highly educated workers will in this case be lower than their marginal social product and some of the increase in aggregate production will go to the group of workers who already were in the city before the human capital inflow. The inflow of human capital therefore increases average wages even when workforce skill-composition is held constant.

¹ But changes in the average level of human capital of the local workforce will have a first-order effect on wages of workers with particular skills if workers with different human capital are imperfect substitutes in production

The theoretical setting that we employ to discuss identification of human capital externalities at the aggregate level is based on the human capital framework used in most theoretical and empirical work at the aggregate level involving human capital.² This framework yields a parsimonious way of capturing imperfect substitutability among workers with many different levels of human capital.³ Moreover, the framework is easily extended to allow for human capital externalities and encompasses the Mincerian approach to externalities.⁴ Our main theoretical result regarding identification of human capital externalities at the aggregate level is that the partial elasticity of average wages with respect to average human capital is equal to the strength of average human capital externalities when workforce skill-composition weights are held constant. This result, which holds whether workers with different human capital are perfect or imperfect substitutes in production, is the basis of what we call the *constant-composition* approach to the identification of human capital externalities at the aggregate level. We also analyze second-order effects of average human capital on average wages holding workforce skill-composition constant. Furthermore, we show that the constant-composition approach can be used to identify human capital externalities that are biased towards workers with high or low levels of human capital.

As an application of the constant-composition approach, we assess the strength of average-schooling externalities in 163 US cities between 1970 and 1990 using instrumental-variable estimation methods to account for endogenous changes in schooling. Our results yield no evidence of significant average-schooling externalities. This finding depends critically on the constant-composition approach being flexible enough for workers with different human capital to be imperfect substitutes in production. Imposing perfect substitutability and using the Mincerian approach with the *same* data and instruments yields

² E.g. Lucas (1988), Mankiw, Romer, and Weil (1992), Benhabib and Spiegel (1994), Klenow and Rodriguez-Clare (1997), Hall and Jones (1999), Topel (1999), Bils and Klenow (2001), de la Fuente and Domenech (2001), Krueger and Lindahl (2001), and Temple (2001).

³ The framework used by Katz and Murphy (1992) captures imperfect substitutability among workers with many different levels of education in an equally parsimonious way. The two frameworks are closely related and our identification results carry over directly to the KM framework. The proofs are only a matter of appropriate relabeling. The KM framework does not encompass the Mincerian approach to human capital externalities however.

⁴ Another framework is the constant-elasticity-of-substitution model used by Moretti (2000). The two main drawbacks of this framework are the assumption that the elasticity of substitution between workers with different schooling is the same whether schooling is very similar or very different and that it is unclear how the strength of average-schooling externalities can be identified.

that a one-year increase in average schooling has a (statistically significant) external effect on productivity of at least 7 percent.

It is well known that wages are in part determined by skills that are unobservable to empirical researchers. This raises the question of how our approach to average-schooling externalities is affected by workers with high wages due to unobservable characteristics (e.g. ability) moving into cities that experience rapidly increasing levels of average schooling. Our data allow us to distinguish between individuals who have worked for a longer period in a city and individuals who moved into the city recently. Hence, we can estimate the wage-differential between “movers” and “stayers” conditional on observable characteristics like education and experience. Our empirical results indicate that this mover-stayer wage-differential is positively correlated with the increase in average schooling in cities between 1970 and 1990 for most education levels. This provides support for the view that cities experiencing rapidly increasing levels of average schooling attract workers with higher ability and that least-squares estimates of average-schooling externalities at the city-level may be biased upwards for this reason. Mover-stayer wage-differentials are not significantly correlated with the increase in average schooling predicted by our instruments for most education levels however. The only exception is the wage-differential for the group of workers with 9 to 12 years of schooling, which is positively correlated with the predicted increase in average schooling. This suggests that our instrumental-variable estimates may understate the strength of average-schooling externalities but are unlikely to overstate them.

The variables used as instruments for the change in average schooling between 1970 and 1990 are the city-level demographic structure of the workforce and population as well as the population-share of African-Americans, all in 1970, and various interaction terms. These variables have predictive power for the change in average schooling at the city-level because younger individuals entering the labor force during this time period had higher levels of schooling than workers going into retirement and because African-Americans were rapidly catching-up in schooling levels with the rest of the population. Our identifying hypothesis is that the variables used as instruments affect aggregate productivity growth between 1970 and 1990 at the city-level only through the change in schooling and other explanatory variables included in the estimating equation. We check this hypothesis by testing the implied overidentifying restrictions (as well as by including selected instruments directly into the estimating equation) and find it cannot be rejected at standard significance levels. As a further check on the instruments, we use them to estimate the strength of average-schooling externalities between 1970 and 1990 at the US *state-level* and compare

the result to estimates obtained with the state-level compulsory-schooling and child-labor-law instruments used by Acemoglu and Angrist (2000). The two sets of instruments yield basically identical estimates.

Mincerian wage regressions to estimate average-schooling externalities in cities were introduced by Rauch (1993). Assuming that average schooling across cities is exogenous, he finds statistically significant externalities in a cross-section of 237 cities in 1980. Acemoglu and Angrist (2000) use the Mincerian approach to estimate average-schooling externalities at the US state-level, accounting for state-specific fixed effects and endogeneity of average and individual schooling. Their instrumental-variable approach, which exploits differences in compulsory-schooling and child-labor laws across states and over time, yields no evidence for significant average-schooling externalities. Rudd (2000) also implements the Mincerian approach at the US state-level, allowing for state-specific fixed effects and controlling for a variety of state-level variables that may affect wages, and like Acemoglu and Angrist does not find evidence for average-schooling externalities. Moretti (2000) accounts for endogenous supply when estimating externalities to the share of college-educated workers in US cities. His theoretical framework allows for imperfect substitutability of workers with different human capital in production but cannot be used to estimate the strength of externalities.⁵ Conley, Flier, and Tsang (1999) report significant instrumental-variable estimates of aggregate human capital externalities in Malaysian regions. The main difference between these papers and our work is that we show how the strength of average human capital externalities can be identified when workers with different human capital are imperfect substitutes in production. There appears to be no previous work on the identification of biased human capital externalities.

The remainder of the paper is organized in the following way. Section 2 presents the theoretical framework. Section 3 derives the main theoretical results on the identification of human capital externalities at the aggregate level. Section 4 contains the estimating equations and explains the estimation methods used. Section 5 describes the data and discusses the instruments. Section 6 presents our empirical results and section 7 summarizes.

⁵ Moretti argues that his framework yields qualitative evidence for externalities however. The argument is based on his finding that wages of college-educated workers at the city-level increase with their share in the workforce. This result is difficult to interpret however as he ignores aggregate scale effects and does not control for the relative supply of other, possibly complementary, types of workers.

2 The Human Capital Framework with Externalities

We now present the aggregate human capital framework with externalities and discuss the identification problem raised by imperfect substitutability between workers with different levels of human capital. The model is the simplest version of the framework that allows us to discuss this identification problem. Extensions will be discussed later. The geographical units of analysis are taken to be cities.

2.1 Model

Assume that output Y of each city depends on the aggregate amount of labor L and human capital H employed in the city according to the following production function

$$Y = AF(L, H), \quad (1)$$

where A denotes the level of total factor productivity (TFP) in the city and H is

$$H \equiv \int_0^{\infty} xL(x)dx, \quad (2)$$

where $L(x)$ is the number of workers with human capital x in the city (using this notation the aggregate amount of labor in the city is $L \equiv \int L(x)dx$). Assume also that the aggregate production function is twice continuously differentiable and subject to constant returns to scale to labor $L(x)$ for all x (or, alternatively, subject to constant returns to scale to L, H) as well as constant or decreasing returns to human capital, $F_{22}(L, H) \leq 0$.

We will allow for the possibility that the marginal social product of workers with above-average (below-average) human capital is greater (smaller) than their equilibrium wage. This is accomplished by assuming that TFP may be increasing in the average level of human capital $h = H/L$ in the city and that this effect takes the form of an externality (e.g. Lucas (1988)). The supply of human capital in the city will therefore affect aggregate production by increasing output conditional on TFP as well as by increasing TFP. Only the former effect will be reflected by wages in our theoretical framework.

To be more precise assume that firms in city c have access to the production function in (1) and that they maximize profits taking the city-time specific levels of TFP as given. Suppose also that product and labor markets are perfectly competitive and that output is tradable. Under these assumptions the equilibrium product wage of workers with human capital x in a city with average human capital h can be written as

$$w(x, h) = w^L(h) + w^H(h)x, \quad (3)$$

where w^L , w^H will be referred to as the price of labor and the price of human capital respectively. The price of labor captures the equilibrium wage of workers without human capital and the price of human capital the wage increase associated with an additional unit of human capital. Both equilibrium prices are linked to TFP, the supply of labor, and the supply of human capital in the city by the usual marginal productivity conditions

$$w^L(h) = AF_1(1, h), \quad (4)$$

$$w^H(h) = AF_2(1, h), \quad (5)$$

where we have used constant returns to scale of the production function given TFP.

Equation (3) implies that identical workers in the same city earn the same product wage. Identical workers in different cities may earn different product wages in equilibrium however as we assume that cities differ in characteristics that are relevant for workers' utility. Examples of such characteristics are the cost of housing, the quality of public schools, local tax-rates, the degree of air pollution, the crime rate, climate, outdoor recreational opportunities, proximity to family or friends, and the variety or quality of restaurants or sports teams. In this case, competitive labor markets imply that product wages across cities satisfy $U^i(w_c(x), z_c) = U^i(w_k(x), z_k)$ for $i \in I$ and all c, k , where z_c, z_k denote vectors of all characteristics of cities c, k that are relevant for utility U^i of type i workers; i captures heterogeneity in preferences.

The specification used for TFP at the city-level is

$$A = Dh^q L^d, \quad (6)$$

where D stands for exogenous city-time specific factors affecting TFP and h, L are the average level of human capital and aggregate employment in the city. The strength of average human capital externalities is captured by the elasticity q . Aggregate scale effects, where scale is measured by aggregate employment, are captured by d . Aggregate scale as a determinant of productivity at the local geographic level is emphasized in Marshall (1890) and Henderson (1988), and estimated in Sveikauskas (1975), Moomaw (1981), Henderson (1986, 1988), Rauch (1993), and Ciccone and Hall (1996) for example. The specification in (6) implies that the effect of an increase in the aggregate stock of human capital on TFP depends in general on how much of the increase is due to aggregate employment growth and how much is due to the growth of average human capital.

The model presented so far is the simplest framework that allows us to discuss identification of human capital externalities when different levels of human capital may be

imperfect substitutes. It can be extended in several dimensions without affecting our theoretical results on identification or our empirical approach. The most important extension would include physical capital and land as factors of production. Extending the theoretical and empirical work to allow for physical capital is simple when physical capital moves to equalize its rate of return across the geographic units of analysis. Identifying human capital externalities at the aggregate level when physical capital is not perfectly mobile across the geographic units of analysis is also straightforward (this is the relevant case for human capital externalities at the country-level). The main insight of extending the theoretical analysis to allow for land as a factor of production is that the strength of externalities will be identified net of congestion effects. All these extensions are discussed in detail in the appendix. It may be worthwhile to point out that the model with land and physical capital has many similarities with the theoretical work of Roback (1982). The model can also be extended to allow for non-tradable goods and for “pecuniary” externalities due to imperfect competition and increasing returns in the production of non-tradables (e.g. Krugman (1992)). These extensions can also be found in the appendix. The main conclusion of the first extension is that the constant-composition approach identifies externalities in the tradable-goods sector only. The main conclusion of the second extension is that the constant-composition approach can also be used to identify “pecuniary” externalities.

2.2 Substitutability and Returns to Human Capital

The framework described so far is flexible enough to allow workers with different levels of human capital to be perfect or imperfect substitutes in production. Perfect substitutability is equivalent to constant returns to human capital given TFP. To see this notice that constant (marginal) returns to human capital $F_{22}(1,h)=0$ imply that the production function in (1) simplifies to

$$Y = A(BL + H) \tag{7}$$

where B is a (possibly city-time specific) exogenous variable. Hence, workers with different human capital are perfect substitutes (the marginal rate of substitution between any two different types of workers is independent of the proportion of the two types used in production). Moreover, it is straightforward to show that perfect substitutability between workers with different human capital combined with constant returns to scale given TFP implies constant returns to human capital given TFP. Constant returns to human capital given TFP and perfect substitutability between workers with different human capital are therefore equivalent in the aggregate human capital framework. Constant returns to human

capital given TFP combined with (3) to (5) yields that the equilibrium wage schedule simplifies to $w(x) = AB + Ax$. Wages of workers with a given level of human capital and the return to human capital will therefore be independent of the average level of human capital in the city if TFP is held constant. Hence, all effects of the average level of human capital on the equilibrium wage schedule must arise through TFP and can be interpreted as externalities.

Imperfect substitutability between different types of workers in production on the other hand is equivalent to decreasing (marginal) returns to human capital $F_{22}(1, h) < 0$. To see this in a simple way suppose that the supply of workers with low human capital x^l in a city decreases while the supply of workers with high human capital x^h ($x^h > x^l$) increases so as to keep the total number of workers constant. It can be shown that the implied change in the relative wage of low human capital workers $w(x^l)/w(x^h)$ is proportional to $-F_{22}(1, h)(x^h - x^l)^2$ in this case (this result is derived in detail in the appendix).⁶ Hence, the decrease in the supply of low human capital workers and increase in the supply of high human capital workers will increase the relative wage of low human capital workers if and only if there are decreasing returns to human capital. Moreover, the implied increase in the relative wage is smaller the closer x^l to x^h . This is because the closer the levels of human capital of the two types of workers, the better they substitute for one another.

3 Identification of Human Capital Externalities

We first discuss the Mincerian approach to human capital externalities. Then we turn to the constant-composition approach.

3.1 The Mincerian Approach

Suppose that aggregate production is subject to constant returns to human capital given TFP and that human capital is essential in production. In this case the aggregate production function in (1) simplifies to $Y = AH$ and the equilibrium wage schedule at the city-level defined in (3)-(5) can be written as $\log w(x) = \log Dh^q L^d + \log x$ using the formulation for endogenous TFP in (6). Assume also that individual human capital is linked to individual schooling s by $x = \exp(gs)$. In this case the equilibrium log-wage schedule becomes

⁶ If there are *only* two types of workers, the production function in (1) implies that the elasticity of substitution between the two types is inversely proportional to $-F_{22}(1, h)(x^h - x^l)^2$.

$$\log w(s) = \log D + \mathbf{d} \log L + \mathbf{q} \log h + \mathbf{g}s . \quad (8)$$

The strength of average-schooling externalities $\mathbf{q}(\partial \log h / \partial S)$ can therefore be estimated as the effect of average schooling in cities on the intercept of an individual (Mincerian) wage regression or, using a city-specific fixed-effects approach, as the effect of changes in average schooling in cities on changes of the intercept. This is the basis of the Mincerian approach to schooling externalities in Rauch (1993), Acemoglu and Angrist (2000), and Rudd (2000). The strength of aggregate scale effects \mathbf{d} can be estimated as the effect of (changes in) log-employment on (changes in) the intercept.

To get a sense for the possible biases of the Mincerian approach when workers with different levels of education are imperfect substitutes in production assume (without loss of generality) that individual levels of human capital are linked to individual schooling s by $x(s) = \exp(g(s))$. Log-linearizing the equilibrium wage schedule in (3) to (5) around the average level of schooling S yields $\log w(s) = \log w(S) + \mathbf{b}(S)g'(S)(s - S)$, where $w(S)$ is the wage of workers with average schooling S and $\mathbf{b}(S) = w^H x(S) / w(S)$ the share of human capital in the wage of workers with average schooling. The log-linearized equilibrium wage schedule can therefore be written as

$$\log w(s) = \log w(S) - RS + Rs . \quad (9)$$

where $R = \mathbf{b}(S)g'(S)$. There are two main differences between (8) and (9). First, the individual return to schooling R may depend on the average level of human capital. Second, making use of (3) to (5), the marginal effect of average schooling on the Mincerian intercept is equal to

$$\mathbf{q} \frac{\partial \log h}{\partial S} + \frac{F_2(1, h)x(S)}{F_1(1, h) + F_2(1, h)x(S)} g'(S) + \frac{F_{12}(1, h) + F_{22}(1, h)x(S)}{F_1(1, h) + F_2(1, h)x(S)} \frac{\partial h}{\partial S} - R - S \frac{\partial R}{\partial S} \quad (10)$$

and may therefore not be equal to average-schooling externalities $\mathbf{q}(\partial \log h / \partial S)$. Using $R = \mathbf{b}(S)g'(S)$, $\mathbf{b}(S) = F_2(1, h)x(S) / (F_1(1, h) + F_2(1, h)x(S))$ and that twice continuous differentiability and constant returns to scale given TFP of the production function in (1) imply $F_{12}(1, h) + F_{22}(1, h)h = F_{21}(1, h) + F_{22}(1, h)h = 0$, we obtain that the difference between the marginal effect of average schooling on the Mincerian intercept and the strength of average-schooling externalities is equal to

$$\text{Bias of Mincerian Approach} = -S \frac{\partial R}{\partial S} + (w(S) - w) \left(\frac{F_{22}(1, h)}{AF_2(1, h)} \right) \frac{\partial h}{\partial S} , \quad (11)$$

where we also used that (3) and (5) imply that the difference between average human capital and the human capital of the workers with average schooling can be written as $h - x(S) = (w - w(S)) / AF_2(1, h)$. Hence, the bias of the Mincerian approach to average-schooling externalities when workers with different levels of human capital are imperfect substitutes depends on two main factors. First, whether the individual return to schooling increases or decreases with the average level of schooling. Second, whether the average wage is greater or smaller than the wage of workers with the average level of schooling.

The main drawback of the Mincerian approach to human capital externalities at the aggregate level is therefore that it yields biased estimates when workers with different human capital are imperfect substitutes. Another drawback is its reliance on Mincerian wage regressions. These regressions raise the concern of endogeneity and mismeasurement of individual schooling and, more generally, correct econometric specification. Some of these problems may be addressed by using instrumental-variable estimation methods (e.g. Acemoglu and Angrist (2000)). But good instruments for individual schooling will usually be unavailable at the local geographic level where externalities are often likely to be strongest. For example, none of the usual instruments for individual schooling is available at the level of US cities.

3.2 The Constant-Composition Approach

The main advantage of the constant-composition approach to the identification of human capital externalities at the aggregate level compared to the Mincerian approach is that it is theoretically valid whether workers with different education are perfect or imperfect substitutes in production. Moreover, the approach does not require estimation of the return to schooling at the individual level. These two results are demonstrated first. Then we turn to second-order effects of average schooling on average wages holding the workforce composition constant and apply the constant-composition approach to the case where human capital externalities may be biased towards workers with high or low human capital.

3.2.1 Basic Constant-Composition Approach

The constant-composition approach is based on the theoretical result that the partial elasticity of average wages with respect to average human capital holding workforce skill-composition weights constant is equal to the strength of average human capital externalities (whether workers with different levels of human capital are perfect or imperfect substitutes). To prove this result notice that the average wage in a city can be written as

$$w(h, l(x) : x \geq 0) = \int_0^{\infty} w(x, h) l(x) dx. \quad (12)$$

This notation emphasizes that average wages depend on individual wages of workers with human capital x as well as workforce skill-composition weights $l(x) = L(x)/L$ and that individual wages depend on average human capital in the city. We can now state our main theoretical result.

Proposition 1: The elasticity of the average wage with respect to the average level of human capital yields the strength of average human capital externalities when workforce skill-composition weights $l(x)$ are held constant

$$\left. \frac{\partial w}{\partial h} \frac{h}{w} \right|_{l(x) \text{ constant } \forall x} = \int_0^{\infty} \left(\frac{l(x)w(x, h)}{w} \right) \left(\frac{\partial w(x, h)}{\partial h} \frac{h}{w(x, h)} \right) dx = \mathbf{q}. \quad (13)$$

Proof: There are different ways to prove this result. One is to notice the relationship between (13) and the dual approach to TFP accounting.⁷ But it is also possible to give a proof that mirrors the intuitive explanation given in the introduction. Suppose that the shares of workers with different human capital go from $\{l(x) : x \geq 0\}$ to $\{l^*(x) : x \geq 0\}$ and that the implied increase in average human capital is Δh . Using the equilibrium wage schedule in (3) and ignoring aggregate scale effects for simplicity yields that the average wage holding workforce skill-composition constant increases by $\int_0^{\infty} (w^L(h + \Delta h) + w^H(h + \Delta h)x)l(x)dx - Dh^q F(1, h)$. The first term can be written as $\int_0^{\infty} (w^L(h + \Delta h) + w^H(h + \Delta h)x)l^*(x)dx + \int_0^{\infty} (w^L(h + \Delta h) + w^H(h + \Delta h)x)(l(x) - l^*(x))dx$, the average wage using the new workforce composition minus the increase in the average wage due to the change in the workforce composition, which simplifies to $D(h + \Delta h)^q F(1, h + \Delta h) - w^H(h + \Delta h)\Delta h$. Hence, the percentage increase in the average wage

⁷ To see this notice that the equilibrium wage schedule in (3) implies that the left-hand side of (13) is equal to $\mathbf{b}e^H + (1 - \mathbf{b})e^L$, where $e^i = (h/w^i)(\partial w^i/\partial h)$ for $i = L, H$ and \mathbf{b} denotes the share of human capital in the average wage, i.e. a weighted average of the elasticities of the price of labor and human capital with respect to average human capital with weights equal to the shares of labor and human capital in the average wage. The proof that this weighted average is equal to the strength of average human capital externalities is very similar to the derivation of the dual approach to TFP accounting. The only difference is that instead of considering the change in TFP associated with the passing of time (dual TFP accounting) we consider the change in TFP associated with an increase in average human capital.

holding workforce skill-composition weights constant, relative to the percentage increase in average human capital, becomes

$$\frac{\frac{D(h + \Delta h)^q F(1, h + \Delta h) - Dh^q F(1, h + \Delta h)}{Dh^q F(1, h)}}{\frac{\Delta h}{h}} + \frac{Dh^q F(1, h + \Delta h) - Dh^q F(1, h) - w^H (h + \Delta h)\Delta h}{Dh^q F(1, h)} \frac{\Delta h}{h}$$

which simplifies to

$$\frac{(h + \Delta h)^q - h^q}{\Delta h} \frac{DF(1, h + \Delta h)}{h^{q-1} DF(1, h)} + \frac{\frac{h^q DF(1, h + \Delta h) - h^q DF(1, h)}{\Delta h} - w^H (h + \Delta h)}{h^{q-1} DF(1, h)}.$$

As the increase in average human capital becomes small, the second term converges to an expression that is proportional to the difference between the marginal product of human capital given TFP and the price of human capital, $Dh^q F_2(1, h) - w^H(h)$, which is zero in equilibrium. The first term converges to q , which is the strength of human capital externalities. **Q.E.D.**

This proposition suggests that we can estimate the strength of average-schooling externalities $q(\partial \log h / \partial S)$ in cities between 1970 and 1990 in two steps. First, obtain the average wage in 1990 in each city using the 1970 workforce composition, $w_{c1990}^F = \sum_Z w_{c1990}(Z) l_{c1970}(Z)$ where Z is the vector of all observable characteristics of workers. Second, estimate the effect of the increase in average schooling in cities 1970-1990, $\Delta S_{c1970-1990} = S_{c1990} - S_{c1970}$, on the log-change in wages holding workforce skill-composition weights constant, $\Delta \log w_{c1970-1990}^F = \log w_{c1990}^F - \log w_{c1970}^F$.

Regarding the identification of aggregate scale effects, it is straightforward to show that the strength of aggregate scale externalities d is equal to the partial elasticity of average wages (whether the workforce composition is held constant or not) with respect to aggregate employment.

3.2.2 Second-Order Effects

So far we have concentrated on first-order effects of the average level of human capital on average wages holding workforce skill-composition weights constant. We now turn to the analysis of second-order effects. The next proposition proves that second-order effects are always positive.

Proposition 2: Suppose that the aggregate production function in (1) is three times continuously differentiable. Then the second-order effect of the log of average human capital on the log of average wages when holding workforce skill-composition weights constant is

$$\mathbf{s} = - \left. \frac{F_{22}(1, h)h^2}{F(1, h)} \right|_{h=h_0} \geq 0, \quad (14)$$

where $h_0 = \int l(x)xdx$. The quadratic approximation of the relationship between the log-change of average wages holding workforce skill-composition weights constant and the log-change of average human capital is therefore

$$\Delta \log w(h, l(x) : x \geq 0) \Big|_{l(x) \text{ constant } \forall x} = \mathbf{q}(\Delta \log h) + \mathbf{s}(\Delta \log h)^2. \quad (15)$$

Proof: The first order effect is

$$\mathbf{q} + \frac{\partial \log(F_1(1, h) + F_2(1, h)h_0)}{\partial \log h} = \mathbf{q} + \frac{F_{12}(1, h) + F_{22}(1, h)h_0}{F_1(1, h) + F_2(1, h)h_0} h.$$

evaluated at $h = h_0$. The second-order effect can be obtained by differentiating the expression above with respect to $\log h$ and evaluating at $h = h_0$. Differentiation yields

$$h(F_{12}(1, h) + F_{22}(1, h)h_0) \frac{\partial \left(\frac{h}{F_1(1, h) + F_2(1, h)h_0} \right)}{\partial h} + \frac{F_{122}(1, h) + F_{222}(1, h)h_0}{F_1(1, h) + F_2(1, h)h_0} h^2$$

The first term evaluated at $h = h_0$ is zero because constant returns to scale of the production function given TFP implies that $F_2(L, H)$ is homogenous of degree zero, which combined

with twice continuous differentiability of the production function yields $F_{12}(1, h) + F_{22}(1, h)h_0 = F_{21}(1, h) + F_{22}(1, h)h_0 = 0$ for $h = h_0$. To simplify the second term notice that constant returns to scale given TFP also imply that $F_1(1, h_0) + F_2(1, h_0)h_0 = F(1, h_0)$ and that $F_{22}(L, H)$ is homogenous of degree minus one. The latter combined with three times continuous differentiability of the production function yields $-F_{22}(1, h_0) = F_{221}(1, h_0) + F_{222}(1, h_0)h_0 = F_{122}(1, h_0) + F_{222}(1, h_0)h_0$. Hence, the second term evaluated at $h = h_0$ becomes $-F_{22}(1, h_0)(h_0)^2 / F(1, h_0)$. **Q.E.D.**

An immediate implication of (14) is that the second-order effect of average human capital on average wages holding workforce skill-composition weights constant is zero if and only if workers with different human capital are perfect substitutes (returns to human capital in production are constant).

The intuition for this last proposition is simple. Suppose that human capital externalities are absent and that returns to human capital are constant. In this case the marginal social product of human capital does not depend on the average level of human capital used in production. Hence, the price of human capital reflects the marginal social product as well as the intra-marginal social product of human capital. Even a large increase in average human capital will therefore not result in an increase in average wages holding workforce skill-composition constant. In the case human capital externalities are absent and the marginal product of human capital is strictly decreasing in the average level of human capital used in production, however, the price of human capital reflects the marginal social product of human capital but is below the intra-marginal social product of human capital. Hence, a large increase in average human capital will result in an increase in average wages holding workforce skill-composition constant (even if wages reflect marginal social products).

When the production function in (1) is of the constant-elasticity-of-substitution type, with the elasticity of substitution between labor and human capital equal to e , the second-order effect can be written as

$$\mathbf{s} = \frac{\mathbf{b}(1 - \mathbf{b})}{e}, \quad (16)$$

where \mathbf{b} is the share of human capital in the average wage. This result will be useful later when we assess the bias of the basic constant-composition approach to average-schooling externalities when second-order effects are important but omitted in the estimating equation.

3.2.3 Biased Human Capital Externalities

Our analysis so far has maintained that human capital externalities enter production in a Hicks-neutral way. We now turn to the case where human capital externalities at the aggregate level may be biased towards workers with high levels of human capital or workers with low levels of human capital. To do so the aggregate production function in (1) is replaced by

$$Y = F(A_L L, A_H H), \quad (17)$$

where

$$A_L = Dh^{q_L} \text{ and } A_H = Dh^{q_H}; \quad (18)$$

q_L, q_H capture externalities of average human capital at the city-level. For simplicity it is assumed that aggregate scale effects are absent. We also assume that the production function is twice continuously differentiable and subject to constant returns to scale given A_L, A_H as well as constant or decreasing returns to human capital given A_L, A_H , $F_{22}(A_L L, A_H H) \leq 0$. The specification in (17) and (18) implies that human capital externalities affect relative wages of workers with different human capital if $q_L \neq q_H$ and the elasticity of substitution between L and H is different from unity.

To determine the strength of average human capital externalities implied by (17) and (18) suppose that average human capital increases by one percent. The resulting increase in average labor productivity is $q_L(1-b) + (1+q_H)b$ where b is the share of human capital in the average wage. Of this total increase, $q_L(1-b) + q_H b$ is due to human capital externalities and will be referred to as the strength of human capital externalities at the aggregate level. The next proposition states that the strength of human capital externalities when externalities may be biased can be identified with the constant-composition approach.

Proposition 3: Suppose that the aggregate production function is given by (17). Then the elasticity of average wages with respect to average human capital holding workforce skill-composition weights constant is equal to $q_L(1-b) + q_H b$.

Proof: The aggregate production function implies that the equilibrium wage schedule is given by $w(x, h) = w^L + w^H h$ where the equilibrium prices of labor and human capital are given by $w^L = DF_1(h^{q_L} L, h^{q_H} H) h^{q_L}$ and $w^H = DF_2(h^{q_L} L, h^{q_H} H) h^{q_H}$. This equilibrium wage schedule implies

$$\left. \frac{\partial w}{\partial h} \frac{h}{w} \right|_{l(x) \text{ constant} \forall x} = \left. \frac{\partial \log \left(DF_1(h^{q_L} L, h^{q_H} H) h^{q_L} + DF_2(h^{q_L} L, h^{q_H} H) h^{q_H} h_0 \right)}{\partial \log h} \right|_{h=h_0} .$$

Constant returns to scale of the aggregate production function given A_L and A_H yields that the marginal product of human capital is homogenous of degree zero. The right-hand-side of the equation can therefore be written as

$$\begin{aligned} & \left. \frac{\partial \log \left(DF_1(1, h^{1+q_H-q_L}) h^{q_L} + DF_2(1, h^{1+q_H-q_L}) h^{q_H} h_0 \right)}{\partial \log h} \right|_{h=h_0} \\ &= (1+q_H-q_L) h^{q_H-q_L} \left(DF_{12}(1, h^{1+q_H-q_L}) h^{q_L} + DF_{22}(1, h^{1+q_H-q_L}) h^{q_H} h_0 \right) \frac{h}{w} \Big|_{h=h_0} \\ & \quad + \left(\frac{q_L w^L}{h} + \frac{q_H w^H h_0}{h} \right) \frac{h}{w} \Big|_{h=h_0} . \end{aligned}$$

Homogeneity of degree zero of the marginal product of human capital combined with the aggregate production function being twice continuously differentiable implies that $F_{12}(1, h^{1+q_H-q_L}) h^{q_L} + F_{22}(1, h^{1+q_H-q_L}) h^{q_H} h_0 = F_{21}(1, h^{1+q_H-q_L}) h^{q_L} + F_{22}(1, h^{1+q_H-q_L}) h^{q_H} h_0 = 0$ for $h = h_0$. Hence,

$$\left. \frac{\partial w}{\partial h} \frac{h}{w} \right|_{l(x) \text{ constant} \forall x} = q_L (1-b) + q_H b . \quad \text{Q.E.D.}$$

4 Estimation

We first describe estimation of average-schooling externalities in cities using the Mincerian approach and then turn to the constant-composition approach.

4.1 The Mincerian Approach

The first step of the Mincerian approach to average-schooling externalities at the city-level 1970-1990 consists of estimation of the following Mincerian wage regression for 1970 and 1990

$$\log w_{ict}(s, e, G, R, M) = a_{ct} + b_t s_{ict} + c_t(e_{ict}) + \mathbf{f}_{Gqt} G_{ict} + \mathbf{f}_{Rqt} R_{ict} + \mathbf{f}_{Mqt} M_{ict} + v, \quad (19)$$

where t stands for either 1970 or 1990; ic denotes individual i in city c ; s, e are individual schooling and potential experience (age minus years of schooling minus six); G, R, M are dummies for gender, race, and marital status; a is a city-specific intercept; b is the individual return to schooling; $c(e)$ is a quadratic function in potential experience; and v captures the variation in log-wages not explained by the right-hand-side variables. The specification allows the effect of gender, race, and marital status on wages to differ across five macro-regions (South, East, Midwest, Mountain, and West) indexed by q . We will estimate (19) using least squares.⁸

The second step of the Mincerian approach consists of estimating the effect of changes in average schooling on changes in the estimated city-specific intercept \hat{a}_{ct} of the Mincerian wage regression using

$$\Delta \hat{a}_{c70-90} = \hat{a}_{c1990} - \hat{a}_{c1970} = \text{Controls} + \mathbf{d} \Delta \log L_{c70-90} + \mathbf{a} \Delta S_{c70-90} + u; \quad (20)$$

the control variables used are a constant combined with four (of the five) macro-region dummies and the change in average potential experience across cities between 1970 and 1990; $\Delta \log L$, ΔS are the change in log-employment and average schooling across cities between 1970 and 1990; and u captures the variation in the intercept not explained by the right-hand-side variables. Notice that city-specific fixed effects do not affect our empirical analysis because average-schooling externalities are estimated using changes over time.

Equation (20) will be estimated using two-stage least squares (2SLS) with the following instruments: individuals in the city younger than 18 per adult in 1970 (YOUNG70); YOUNG70 squared; the share of the city-workforce older than 50 in 1970; the share of African-Americans in the city-population in 1970; two interaction terms; and the constant combined with four (of the five) macro-region dummies (the instruments will be discussed in more detail in the next section). Our identifying hypothesis is that these variables affect aggregate productivity growth between 1970 and 1990 at the city-level only through the change in schooling and the other explanatory variables included in the estimating equation. Given that the number of instruments exceeds the number of right-hand-side variables, we can test the resulting overidentifying restrictions.

⁸ Acemoglu and Angrist (2000) estimate average-schooling externalities at the US state-level using the Mincerian approach with and without instruments for individual schooling and find no difference between the two approaches.

We also want to see whether changes in average schooling over time affect the individual return to schooling at the city-level. To do so we first re-estimate (19) for 1970 and 1990 allowing the effect of individual schooling on individual wages to vary across cities and then relate the change in the estimated return to schooling over time $\Delta \hat{b}_{c70-90} = \hat{b}_{c1990} - \hat{b}_{c1970}$ to the change in average schooling and the log-scale of production using

$$\Delta \hat{b}_{c70-90} = \text{Controls} + r\Delta \log L_{c70-90} + m\Delta S_{c70-90} + u. \quad (21)$$

The controls, method of estimation, and instruments used are the same as in (20). If the assumption underlying the Mincerian approach to average-schooling externalities were satisfied then the effect of average schooling on individual return m should be insignificantly different from zero.

4.2 The Constant-Composition Approach

We first present the estimating equations of the basic constant-composition approach. Then we discuss second-order effects and introduce a method to evaluate how the constant-composition approach is affected by selective migration of workers with high wages due to unobservable characteristics.

4.2.1 Basic Specification

The constant-composition approach consists of estimating the effect of changes in average schooling over time on the log-change in average wages holding workforce skill-composition weights constant. The change in log-wages holding the workforce composition constant is constructed by first estimating wage regressions for 1970 and 1990 that relate the log-wage of individuals with levels of schooling and potential experience s, e to city-specific effects $\log w_{ct}(s, e)$ and dummies for gender, race, and marital status

$$\log w_{ict}(s, e; G, R, M) = \log w_{ct}(s, e) + \mathbf{I}_{Gqt} G_{ict} + \mathbf{I}_{Rqt} R_{ict} + \mathbf{I}_{Mqt} M_{ict} + v. \quad (22)$$

The regression is set up so that the intercept corresponds to married white males. The estimation method used is least squares.

Estimating (22) yields city-specific average wages of workers with levels of schooling and potential experience s, e adjusted for region-specific gender, marital status, and race wage-differentials, $\hat{w}_{ct}(s, e)$, for 1970 and 1990. Our measure of average wages holding workforce skill-composition constant is based on these adjusted wages

$$\hat{w}_{ct}^{AF} = \sum_{s,e} \hat{w}_{ct}(s,e) l_{c1970}(s,e), \quad (23)$$

where $l_{c1970}(s,e)$ is the fraction of workers with individual schooling and potential experience s,e in city c in 1970.

Estimation of the strength of average-schooling externalities in cities between 1970 and 1990 is based on adjusted average wages holding workforce skill-composition weights constant. In particular, we estimate

$$\begin{aligned} \Delta \log \hat{w}_{c,70-90}^{AF} &= \log \hat{w}_{c1990,1970}^{AF} - \log \hat{w}_{c1970,1970}^{AF} \\ &= \text{Controls} + \mathbf{d} \Delta \log L_{c,70-90} + \mathbf{a} \Delta S_{c,70-90} + u, \end{aligned} \quad (24)$$

using the same controls, method of estimation, and instruments as in (20).

4.2.2 Second-Order Effects

The estimating equation in (24) corresponds to the basic constant-composition approach to human capital externalities. It may seem straightforward to extend this approach to account for the second-order effects in (15) by adding the change in average schooling squared as a right-hand-side variable. This is difficult in practice however because changes in average schooling and changes in average schooling squared across cities are almost perfectly correlated in our data (the simple correlation coefficient 1970-1990 is 0.94).⁹ As a result our instruments do not predict any of the variation in average schooling squared once the change in average schooling is accounted for and the strength of second-order effects cannot be estimated in practice.

It is however possible to get a sense for the *sign* of the bias of the basic constant-composition approach to human capital externalities should (omitted) second-order effects be important. To see how assume that the log-change in average human capital is proportional to the change in average schooling. Suppose also that the aggregate production function is of the constant-elasticity-of-substitution type with $\mathbf{e} > 0$. In this case (15) and (16) imply that the second-order effect is proportional to $\mathbf{b}_c(1-\mathbf{b}_c)(\Delta S_c)^2$. Hence, the bias of the basic constant-composition approach should second-order effects be important is proportional to the coefficient on ΔS_c when equation (24) is estimated using $\mathbf{b}_c(1-\mathbf{b}_c)(\Delta S_c)^2$ as the left-hand-side variable (e.g. Greene (1999)). Signing the bias can be

⁹ Another difficulty is that the strength of the second-order effect \mathbf{s} will generally depend on the average level of human capital and would therefore be a city-time specific parameter.

simplified further if we assume that the production function is Cobb-Douglas with identical shares of human capital in the average wage across cities. In this case the bias is proportional to the coefficient on ΔS_c when equation (24) is estimated using $(\Delta S_c)^2$ as the left-hand-side variable.

Signing the bias when the share of human capital in the average wage is not constant across cities requires estimating $\mathbf{b}_c(1-\mathbf{b}_c)$. To see how this can be done notice that (9) implies that the share of human capital in the wage of workers with average schooling $\mathbf{b}_c(S)$ is linked to the individual return to schooling in a city-level wage regression b_c by $\mathbf{b}_c(S_c)\mathbf{g}_c = b_c$ where $\mathbf{g}_c = g(S_c)$. Furthermore, using the definition of \mathbf{b}_c and $\mathbf{b}_c(S)$ yields that the share of human capital in the wage of workers with average schooling is linked to the share of human capital in the average wage by $\mathbf{b}_c = \mathbf{b}_c(S_c)/[(1-\mathbf{b}_c(S_c))\mathbf{r}_c + \mathbf{b}_c(S_c)]$ where $\mathbf{r}_c = x_c(S_c)/h_c \geq 0$ is the human capital of workers with average schooling relative to average human capital. Suppose that both the elasticity of human capital with respect to schooling \mathbf{g}_c and the human capital of workers with average schooling relative to average human capital \mathbf{r}_c are approximately constant across US cities, $\mathbf{g}_c = \mathbf{g}$ and $\mathbf{r}_c = \mathbf{r}$. Then the share of human capital in the average wage is linked to the individual return to schooling by

$$\mathbf{b}_c = \frac{b_c}{\mathbf{r}\mathbf{g} + b_c(1-\mathbf{r})}. \quad (25)$$

The share of human capital in the average wage must be smaller or equal to unity, $\mathbf{b}_c \leq 1$, which implies that $\mathbf{g} \geq \max_c b_c$. Moreover, assume that the share of human capital in the average wage is at least one percent, $\mathbf{b}_c \geq 0.01$. This yields that $\mathbf{g} \leq [100/\mathbf{r} - (1-\mathbf{r})/\mathbf{r}]\min_c b_c$. Finally, individual wages in the US are a convex function of individual schooling. This implies that the average wage is above the wage of workers with average schooling and, combined with (3), that average human capital is greater than the human capital of workers with average schooling, $\mathbf{r} \leq 1$. Hence, \mathbf{g}, \mathbf{r} must satisfy $\max_c b_c \leq \mathbf{g} \leq [100/\mathbf{r} - (1-\mathbf{r})/\mathbf{r}]\min_c b_c$ and $0 \leq \mathbf{r} \leq 1$. We can therefore use estimates of the individual return to schooling across cities \hat{b}_c and (25) to estimate $\mathbf{b}_c(1-\mathbf{b}_c)$, and hence the bias of the basic constant-composition approach should (omitted) second-order effects be important, for all values of \mathbf{g}, \mathbf{r} satisfying these restrictions.

4.2.3 Unobserved Heterogeneity

The constant-composition approach to identifying human capital externalities requires holding workforce skill-composition weights used to calculate the change in the log of average wages constant. The estimating equation in (24) only fixes the composition of

observable characteristics however. Empirical results based on this equation could therefore be biased because of changes in the composition of unobservable skills due to selective migration (across cities). For example, the increase in average schooling at the city-level may be positively correlated with the increase in the average ability of the workforce. Least-squares estimates of the strength of average-schooling externalities would in this case be biased upwards while the bias of 2SLS estimates would depend on the correlation between the change in average ability and the change in average schooling predicted by the instruments.

To explore how our estimates of the strength of average-schooling externalities may be affected by selective migration we first estimate the wage-differential conditional on observable characteristics between workers who moved into the city recently (“movers”) and workers who lived in the city for a longer period (“stayers”). This can be done for both 1980 and 1990 using two different samples, which we refer to as samples A and B. Sample A includes all workers who five years before we observe their wages did not live in the same city plus all workers who lived in the same house since 1970. Sample B includes all workers who did not live in the same city five years before we observe their wages plus all workers who already lived in the same city five years before and were born in the state where they reside. For both samples we estimate the following wage regression given individual schooling s in 1980 and 1990,

$$\log w_{ict}(s; e, G, R, M) = a_{cst} + c_{st}(e_{ict}) + \mathbf{f}_{cst} MOVER_{ict} + \mathbf{f}_{sGqt} G_{ict} + \mathbf{f}_{sRqt} R_{ict} + \mathbf{f}_{sMqt} M_{ict}. \quad (26)$$

The dummy $MOVER$ is equal to unity if the worker did not live in the city where we observe her in 1980 or 1990 five years before. The definition of stayer in sample B is therefore much wider than in sample A and includes workers who moved into the city after 1970 as long as they did so more than five years before we observe their wages and were born in the state of residence. The estimated city-specific mover-stayer wage-differential for workers with schooling s , $\hat{\mathbf{f}}_{cst}$, for 1980 and 1990 is then used as left-hand-side variable in (24). Estimating this equation using both least squares and 2SLS with the usual instruments gives us a sense whether unobservable characteristics translating into high wages are positively correlated with the increase in average schooling or the increase in average schooling predicted by our instruments.

5 Data and Instruments

We use data on approximately 4 million individuals in 163 cities in 1970, 1980, and 1990. The data comes from the public use micro samples of the US Census (US Bureau of Census (1970, 1980, 1990)). Individual wages are measured per hour worked. When we implement the individual wage regression in (19), we use years of potential experience and eleven levels of schooling. When we implement (22), we partition potential experience in five intervals and schooling in seven intervals, which yields a total of thirty-five schooling-experience combinations. Race consists of dummies for: White; Black; Hispanic; Indian or Eskimo; Japanese, Chinese, or Filipino; and Pacific Islander or Hawaiian. Details on the data can be found in the appendix.

Our definition of cities corresponds with some exceptions to the US Bureau of Census definition of standard metropolitan statistical areas (SMSAs) in 1990 and is explained in detail in the appendix. City-level employment in 1970 and 1990 is obtained by summing employment of all counties that were contained in the city in 1990. County-employment is the number of people with part-time or full-time jobs and comes from the U.S. Department of Commerce (US Department of Commerce (1992)). We only consider employment in the private sector and exclude agriculture and mining.

Average years of schooling and experience at the city-level are obtained by aggregating years of schooling and potential experience of individuals in the city. Average schooling across cities rose by 1.12 years during the 20-year period 1970-1990. The standard deviation of the increase in average schooling was 0.56 and the maximal increase 2.1 years. Average potential experience across cities fell by 5.3 years.

Table 1 contains the results of regressing the 1970-1990 increase in average schooling and average experience across cities on the 1970 instruments using the specification that fits the data best. The R^2 of the average schooling regression is 48 percent without macro-region dummies and 57 percent with macro-region dummies.¹⁰ The R^2 of the average-experience regression is 47 percent without macro-region dummies and 51 percent with macro-region dummies. The coefficient estimates of the average-schooling regressions in columns (1) and (2) combined with the sample values of the explanatory variables yield the

¹⁰ The instruments (without macro-region dummies) explain 38 percent of the increase in the share of workers with a high school education or more, 31 percent of the increase in the share of workers with some college or more, 32 percent of the increase in the share of workers with a college education or more, 25 percent of the increase in the share of workers with a high school education only, and 37 percent of the decrease in the share of high school dropouts.

following three main results (the non-linear specification implies that coefficient estimates must be combined with the sample values of the explanatory variables to assess the effect of changes in the explanatory variables on average schooling). First, cities with a larger share of workers older than 50 in 1970 (AGE50P70) experienced a greater increase in average schooling between 1970 and 1990. This is because workers who retired in this period had levels of education below the workforce average. Second, cities with a larger number of people younger than 18 per adult in 1970 (YOUNG70) experienced a greater increase in average schooling between 1970 and 1990. This is because young people entering the workforce in this period had levels of education above the workforce average. The quadratic specification implies that the marginal effect of YOUNG70 on the increase in average schooling was larger in cities with a larger number of people younger than 18 per adult in 1970 (and also that the marginal effect would be negative for small values of YOUNG70; for sample values the effect is always positive however). When we add macro-region dummies in column (2), YOUNG70 and YOUNG70 squared are no longer individually significant but remain jointly significant at the 5-percent level. Third, cities with a larger population share of African-Americans in 1970 experienced a greater increase in average schooling between 1970 and 1990. This is because African-Americans were catching up rapidly to average levels of education over this time-period. The coefficient estimates of the average-experience regressions in columns (3) and (4) imply that, for sample values of the explanatory variables, a larger number of people younger than 18 per adult in 1970 and a larger share of workers older than 50 in 1970 was associated with a larger decrease in average experience between 1970 and 1990.

6 Results

We first discuss the results using the constant-composition approach to average-schooling externalities and then compare the constant-composition results with those of the Mincerian approach.

6.1 The Constant-Composition Approach

After presenting the results of the basic constant-composition approach, we discuss the empirical implications of second-order effects and selective migration of workers with high wages due to unobservable characteristics.

6.1.1 Basic Specification

Table 2 contains the results of estimating (24) with 2SLS and the instruments discussed in the previous section.¹¹ Column (1) uses the constant and four (of the five) macro-region dummies as controls. The estimate of the strength of average-schooling externalities is 0.014 with a standard error of 0.03 and hence highly insignificant. Column (2) eliminates the (individually and jointly) insignificant macro-region dummies SOUTH and WEST. The estimate of the strength of average-schooling externalities is now -0.004 with a standard error of 0.017. Column (3) uses the constant and four (of the five) macro-region dummies as well as the change in average potential experience 1970-1990 as controls. The estimate of the strength of average-schooling externalities does not change much compared to the specification without average experience in column (1). Changes in average potential experience have a significantly negative effect on average wages holding workforce skill-composition constant, which means that cities where the average age of the workforce predicted by our instruments fell more than average saw an above-average increase of average wages holding workforce skill-composition constant. This suggests that these cities experienced an inflow of workers with high wages due to unobservable characteristics.¹² The P-value of the test of overidentifying restrictions in the last row (0.53) indicates that these restrictions cannot be rejected at standard significance levels. Column (4) eliminates the (individually and jointly) insignificant macro-region dummies SOUTH and WEST. The estimate of average-schooling externalities is now -0.01 with a standard error of 0.018. The P-value of the test of overidentifying restrictions in the last row (0.41) indicates that these restrictions cannot be rejected at standard significance levels. Columns (5) to (10) estimate equation (24) using selected instruments as control variables. The direct effect of the instruments on average wages holding workforce skill-composition constant is in all cases small and statistically insignificant. For example, when adding the population-share of

¹¹ Least-squares (LS) estimation of (24) is likely to yield biased estimates because both the increase in average schooling and log-employment are endogenous and measured with error. Still, in practice LS estimates of the strength of average-schooling externalities are very similar to 2SLS estimates (the difference is at most half a percentage point) and highly insignificant. This suggests that the different biases present in least-squares estimation tend to offset each other in this particular application.

¹² A piece of evidence supporting this interpretation is our finding that the partial correlation (holding the change in average schooling predicted by our instruments constant) between the change in average potential experience predicted by our instruments and the mover-stayer wage-differential is negative.

African-Americans in 1970 as a control variable in column (7), we find that a 5 percentage points increase in this share lowers average wages holding workforce skill-composition constant by only 0.2 percent (the maximum variation in the share of African-Americans across cities in 1970 is 25 percentage points) and that this effect is highly insignificant. Moreover, estimates of the strength of average-schooling externalities in columns (5) to (10) remain close to zero and insignificant.

Table 3 contains the results of estimating (24) when data on white males only is used to construct constant-composition average wages. The method of estimation is 2SLS with the usual instruments. Columns (1) to (5) use data on all white males adjusted for wage-differentials related to marital status (the adjustment is based on (22)). The results are very similar to those obtained using all workers once the (individually and jointly) insignificant macro-region dummies SOUTH and WEST are eliminated. For example the strength of average-schooling externalities in column (2) is -0.001 with a standard error of 0.021. The P-value of the test of overidentifying restrictions in the last row (0.73) indicates that these restrictions cannot be rejected at standard significance levels. Columns (6) and (7) contain the results of estimating (24) using data on white males aged 40-49 only to construct constant-composition average wages (these are the workers used by Acemoglu and Angrist (2000) to estimate average-schooling externalities at the state-level). Once the (individually and jointly) insignificant macro-region dummies SOUTH and WEST are eliminated, the results are similar to those obtained with all white males and all workers.

Estimates of the strength of aggregate scale effects in tables 2 and 3 are very imprecise and larger than the 4 to 10 percent effect reported in much of the literature (e.g. Henderson (1986, 1988), Ciccone and Hall (1996)).¹³ To see whether our results are sensitive with respect to the strength of aggregate scale effects we re-estimate (24) restricting aggregate scale effects to values between 4 and 10 percent. The results are reported in table 4. Estimates of the strength of average-schooling externalities are in all cases close to the values obtained when the strength of aggregate scale effects is estimated.

Table 5 contains constant-composition-approach estimates of average-schooling externalities between 1970 and 1990 at the US *state-level*. The method of estimation is 2SLS using either the instruments of Acemoglu and Angrist (2000) or our instruments at the state-level. The AA instruments are based on whether state-level compulsory-schooling (child-labor) laws when workers were 14 required 8, 9, 10, or 11 years of minimum

¹³ This is because our instruments predict only 22 percent of the variation of the log-change in aggregate employment 1970-1990 across cities.

schooling (6, 7, 8, or 9 years of schooling before work was permitted). AA construct these instruments using both the state of residence and the state of birth of workers and find that empirical results are very similar in the two cases. Column (1) contains the result of estimating $\Delta \log \hat{w}_{state,70-90}^F = \text{Constant} + \mathbf{a}\Delta S_{state,70-90}$, where state-level constant-composition average wages are constructed using data on white males aged 40-49 only, with the AA instruments (following AA, we do not include the change in log-employment or other control variables in the estimating equation). In particular, the instruments used are the change between 1970 and 1990 of the fraction of workers for whom state-of-residence compulsory-schooling (child-labor) laws when they were 14 required 8, 9, 10, or 11 years of minimum schooling (6, 7, 8, or 9 years of schooling before work is permitted). These instruments combined predict 45 percent of the change in average schooling 1970-1990 (not in table). Column (2) contains the result of estimating the same equation using 2SLS with our instruments at the state-level. Our instruments predict 51 percent of the change in average schooling 1970-1990 at the state-level (not in table). It can be seen that estimates of the strength of average-schooling externalities obtained with both sets of instruments are nearly identical. (The results in table 5 are not comparable to those in AA because they use the Mincerian approach and never consider the 1970-1990 period by itself.¹⁴)

6.1.2 Second-Order Effects

Regressing $(\Delta S_c)^2$ on the right-hand-side variables of (24) using 2SLS and the usual instruments yields a coefficient on ΔS_c of 2.71 with a standard error of 0.14. Figures (1a) to (1c) deal with the sign of the bias when it cannot be assumed that the share of human capital in the average wage is identical across cities. The figures display the estimates of the coefficient on ΔS_c (and the 95-percent confidence bands) when regressing $\hat{b}_c(1-\hat{b}_c)(\Delta S_c)^2$ on the right-hand-side variables of (24) using 2SLS and the usual instruments. The values of $\hat{b}_c(1-\hat{b}_c)$ underlying (1a) are based on (25) assuming that $\mathbf{r}=0.99$ and that \mathbf{g} satisfies $\max_c \hat{b}_c \leq \mathbf{g} \leq [100/\mathbf{r} - (1-\mathbf{r})/\mathbf{r}] \min_c \hat{b}_c$, where \hat{b}_c is the 1970 return to schooling in cities

¹⁴ To reproduce the AA results we also estimate $\Delta \hat{a}_{state}^{AA} = \text{Constant} + \mathbf{a}\Delta S_{state}$ for 1960-1980, where $\Delta \hat{a}_{state}^{AA}$ is the 1960-1980 change in the estimated intercept of a Mincerian least-squares wage regression at the state-level using data on white males 40-49 only. 2SLS yields an estimate of the average-schooling externalities equal to 0.005 with a standard error of 0.029 when using as instruments the 1960-1980 change in the fraction of workers for whom state-of-residence compulsory-schooling (child-labor) laws when they were 14 required 8, 9, 10, or 11 years of minimum schooling (6, 7, 8, or 9 years of schooling before work is permitted). This estimate is almost identical to the result of AA for the same time period.

estimated using (19) with a city-specific individual return to schooling. Figure (1b) and (1c) repeat the same exercise for values of $r=0.5$ and $r=0.01$ respectively. It can be seen from the figures that the bias is significantly positive in all three cases. Repeating the analysis for all values of r between 0.01 and 0.99 in one-percent steps yields that the bias of the basic constant-composition approach to average-schooling externalities in cities should (omitted) second-order effects be important is significantly positive in all cases.

6.1.3 Unobserved Heterogeneity

Table 6 contains the coefficient on ΔS_c when regressing the estimated city-specific mover-stayer wage-differential \hat{f}_{cst} on the right-hand-side variables of (24) using least squares as well as 2SLS with the usual instruments. \hat{f}_{cst} is estimated using (26) for the six individual schooling intervals indicated in the table. The least-squares results indicate a significantly positive partial correlation between the mover-stayer wage-differential and the increase in average schooling at the city-level for workers with 9 to 17 years of schooling in 1980 and workers with 9 to 14 years of schooling in 1990. This finding supports the view that cities with a larger increase in average schooling attract workers with higher wages due to unobservable characteristics. Interestingly, the results are very similar whether stayers (“non-movers”) are defined as workers who lived in the same house since 1970 or workers who lived in the same city five years before we observe their wages and were born in the state where they reside. The only negative partial correlation between the mover-stayer wage-differential and the increase in average schooling can be found for workers with 0 to 9 years of schooling and workers with 17 or more years of schooling in 1980. These estimates are statistically insignificant however. Moreover, workers with 0 to 9 years of schooling are a small fraction of the workforce in both 1980 (6.6 percent) and 1990 (4.3 percent). Similarly workers with 17 or more years of schooling in 1980 are only 10 percent of the workforce.

Table 6 also summarizes the results of regressing \hat{f}_{cst} on the right-hand-side variables of (24) using 2SLS with the usual instruments. It can be seen that there is no significant relationship between the increase in average schooling and the mover-stayer wage-differential with the exception of the significantly positive relationship for workers with 12 to 14 years of schooling in 1990. Furthermore, the estimates are all positive with the exception of workers with 0 to 9 years of schooling in 1990. These workers are only 4.3 percent of the workforce however. The 2SLS results therefore suggest that mover-stayer wage-differentials are not as highly correlated with the increase in average schooling

predicted by our instruments as with the increase in average schooling. Moreover, if our 2SLS estimates are biased because of unobserved heterogeneity and selective migration, they are biased upwards.

6.2 The Mincerian Approach

Table 7 contains estimates of average-schooling externalities at the city-level using the Mincerian approach in (20). The control variables, estimation method, and instruments are the same ones used to implement the constant-composition approach. Columns (1) to (3) contain the results when the underlying Mincerian wage regression in (19) is estimated using all workers. According to columns (1) and (2), the external effect on productivity of an additional year of average schooling is above 7 percent and statistically significant at the 5-percent level. Columns (1) and (2) differ in that the latter eliminates the (individually and jointly) insignificant SOUTH and WEST macro-region dummies. Column (3) contains the average-schooling-externality estimate when the Mincerian intercept in (19) is estimated allowing the individual return to schooling to differ across cities. The estimate triples compared to the two previous columns. Column (4) contains the average-schooling-externality estimate when the Mincerian intercept is estimated using data on white males only (restricting the individual return to schooling to be equal across cities). Results are very similar to the case where all workers are used. The P-value of the test of overidentifying restrictions in the last row of columns (1) to (4) indicates that these restrictions cannot be rejected at standard significance levels.

Table 8 investigates the effect of average schooling in cities on the individual return to schooling. All results are based on estimating (21) using 2SLS with the usual instruments. Column (1) indicates that the change in average schooling between 1970 and 1990 does not have a statistically significant effect on the change in the individual return to schooling when all four macro-region dummies are included in the analysis. The four macro-region dummies are individually and jointly insignificant at standard significance levels however (the P-value of the joint exclusion restriction is 0.41). Columns (2) to (12) estimate the effect of the change in average schooling 1970-1990 on the change in the individual return to schooling when the four macro-region dummies are not used as controls. Column (2) contains the basic specification. The results suggest that a one-year increase in average schooling decreases the individual return to schooling by 1.5 percent (with a standard error of around 0.4 percent). The P-value of the test of overidentifying restrictions in the last row (0.83) indicates that these restrictions cannot be rejected at standard significance levels. Columns

(3) to (6) include selected instruments directly into the estimating equation to see whether they affect the results. The direct effect of the instruments on the change in individual return to schooling is insignificant and the effect of average schooling significantly negative in all cases. Columns (7) to (11) estimate (21) without using the four macro-region dummies as instruments. A one-year increase in average schooling is now estimated to reduce the individual return to schooling by around 1.1 percent (with a standard error of around 0.5 percent). Finally, column (12) estimates the effect of average schooling on the individual return to schooling using data on white males only to estimate the individual return to schooling. The results are very similar to the case where all workers are used.

7 Summary

When workers with different human capital are imperfect substitutes in production, an increase in average human capital changes wages of workers with particular skills but has no first-order effect on average wages holding workforce skill-composition constant if wages reflect marginal social products. Based on this result, we propose an approach to identify human capital externalities at the aggregate level that yields positive externalities if and only if the marginal social product of workers with above-average human capital exceeds their wage. The approach can also be used to identify human capital externalities that are biased towards workers with high or low levels of human capital.

References

- Acemoglu, Daron and Joshua Angrist (2000) "How Large are the Social Returns to Education: Evidence from Compulsory Schooling Laws," in Ben S. Bernanke and Kenneth Rogoff eds, *NBER Macroeconomic Annual 2000*, pp 9-59.
- Autor, David, Larry Katz and Alan Krueger (1998) "Computing Inequality: have Computers changed the Labor Market?" *Quarterly Journal of Economics* 113 (4), pp. 1169-1213.
- Azariadis, Costas and Allen Drazen (1990) "Threshold Externalities in Economic Development," *Quarterly Journal of Economics* 105, pp, 501-26.
- Benhabib, Jess and Mark Spiegel (1994) "The Role of Human Capital in Economic Development, Evidence from Aggregate Cross-Country Data," *Journal of Monetary Economics* 34 (2), pp. 143-173.
- Bils, Mark and Peter Klenow (2000) "Does Schooling Cause Growth?" *American Economic Review*, 90, pp.1160-1193.
- Black, Duncan and Vernon Henderson (1999) "A Theory of Urban Growth," *Journal of Political Economy* 107 (2), pp. 252-284.
- Card, David and Thomas Lemieux (2000) "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort Based Analysis," NBER Working Paper No. W7655, April.
- Ciccone Antonio and Robert Hall (1996) "Productivity and the Density of Economic Activity," *American Economic Review* 86 (1), pp. 54-70.
- Conley, Timothy, Frederick Flier, and Grace Tsiang (1999) "Local Market Human Capital and the Spatial Distribution of Productivity in Malaysia," Mimeo, Northwestern University, February.
- De la Fuente, Angel and Raphael Domenech (2001) "Schooling Data, Technological Diffusion and the Neoclassical Model," *American Economic Review* 91 (2), pp. 323-327.
- Gemmell, Norman (1997) "Externalities to Higher Education," Report to the National Committee of Inquiry into Higher Education, HMSO Copyright Unit, St Clements House, Norwich.
- Greene, William (1999) *Econometric Analysis* 4th Edition, Prentice Hall, New Jersey.
- Hall, Robert and Charles Jones (1999) "Why Do Some Countries Produce So Much More Output Per Worker than Others?" *Quarterly Journal of Economics*, 114(1), pp. 83-116.
- Heckman, James and Peter Klenow (1998) "Human Capital Policy," in Michael Boskin (editor) *Policies to Promote Capital Formation* Hoover Institution.
- Heckman, James (2000) "Policies to Foster Human Capital," *Research in Economics* 51, No 1 March, pp. 3-56.
- Henderson, Vernon (1986) "Efficiency of Resource Usage and City Size," *Journal of Urban Economics* 18, pp. 23-56.
- Henderson, Vernon (1988) *Urban Development: Theory, Fact, and Illusion*, Oxford University Press, Oxford, UK.
- Johnson, George (1997) "Changes in Earnings Inequality: The Role of Demand Shifts," *Journal of Economic Perspectives* 11 (2), pp. 41-54.

- Katz, Larry and Kevin Murphy (1992) "Changes in Relative Wages 1963-1987- Supply and Demand Factors," *Quarterly Journal of Economics* 107 (1), pp. 35-78.
- Klenow, Peter and Andres Rodriguez-Clare (1997) "The Neoclassical Revival in Growth Economics: Has it Gone too Far?" in Ben Bernanke and Julio Rotemberg eds. *NBER Macroeconomic Annual 1997*, pp.73-103, Cambridge MA. MIT Press.
- Krueger, Alan and Mikael Lindahl (2001) "Education for Growth: Why and for Whom?" *Journal of Economic Literature* 39 (4), pp. 1101-1136.
- Krugman, Paul (1992) "Increasing Returns and Economic Geography," *Journal of Political Economy* 99, pp. 483-499.
- Lucas, Robert E. (1988) "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22, pp. 3-41.
- Mankiw, Gregory, David Romer, and David Weil (1992) "A Contribution to the Empirics of Economic Growth" *Quarterly Journal of Economics*, 107 (2), pp. 407-438.
- Marshall, Alfred (1890) *Principles of Economics*, Macmillan, London.
- Moomaw, Ronald (1981), "Productivity and City Size: A Critique of the Evidence," *Quarterly Journal of Economics* 95, pp. 194-223.
- Moretti, Enrico (2000) "Estimating the Social Return to Education: Evidence From Longitudinal and Cross-Sectional Data," Center for Labor Economics, University of California, Berkeley, Working Paper No. 22.
- Rauch, James (1993) "Productivity Gains from Geographic Concentration in Cities," *Journal of Urban Economics* 34, pp. 380-400.
- Roback, Jennifer (1982) "Wages, Rents, and the Quality of Life," *Journal of Political Economy* 90, pp. 1257-1278.
- Rudd, Jeremy (2000) "Empirical Evidence on Human Capital Spillovers," Federal Reserve Board, Finance and Economics Discussion Paper 2000-46, October.
- Sveikauskas, Leo (1975), "The Productivity of Cities," *Quarterly Journal of Economics* 89, pp. 393-413.
- Temple, Jonathan (2001) "Growth effects of education and social capital in the OECD countries," *OECD Economic Studies*, 2001/2, 33, pp.57-101.
- Topel, Robert (1997) "Factor Proportions and Relative Wages," *Journal of Economic Perspectives*, 11(2), pp. 55-74.
- Topel, Robert (1999) "Labor Markets and Economic Growth," in *Handbook of Labor Economics*, ed. Orley Ashenfelter and David Card. Amsterdam: Elsevier Science, pp. 2943-2984.
- U.S. Bureau of Census (1970), *Census of Population and Housing, United States*.
- U.S. Bureau of Census (1980), *Census of Population and Housing, United States*.
- U.S. Bureau of Census (1990), *Census of Population and Housing, United States*.
- US Department of Commerce (1992), Economics and Statistics Administration, Bureau of Economic Analysis, Regional Economic Measurement Division, *REIS CD-ROM*.

Appendix

A.1 Tables and Figures

Table 1: Quality of the 1970 instruments for the change in average schooling and average experience 1970-1990

	Change in average schooling 1970-1990 (ΔS)		Change in average experience 1970-1990 (ΔE)	
	(1)	(2)	(3)	(4)
Share of the city-workforce older than 50 in 1970 (50PLUS70)	5.7** (1.1)	3.6** (1.1)	-25.3** (3.02)	-23.4** (3.34)
Share of African-Americans in the city-population in 1970 (AA70)	11.6** (3.1)	6.8** (2.9)	-3.65 (8.48)	1.25 (8.85)
People in the city younger than 18 per adult in 1970 (YOUNG70)	-3.7* (2.2)	-3.0 (2.2)	12.47** (5.9)	12.79** (6.11)
YOUNG70*YOUNG70	2.7** (1.2)	2.8 (2.1)	-8.1** (3.38)	-8.55** (3.44)
YOUNG70*AA70	-5.9** (2.6)	-2.9 (2.5)	-6.74 (7.38)	-11.18 (7.47)
50PLUS70*AA70	-16.7** (8.2)	-8.8 (8.2)	41.42* (24.45)	39.1* (21.23)
MIDWEST		-0.17** (0.06)		-0.01 (0.2)
MOUNTAIN		-0.29** (0.14)		-0.14 (0.42)
WEST		-0.47** (0.07)		0.52** (0.23)
SOUTH		-0.14** (0.07)		0.05 (0.21)
R^2	0.48	0.57	0.47	0.51

Notes: Results of regressing the increase in average years of schooling and average years of potential experience 1970-1990 at the city-level on a constant and the variables in the leftmost column using least squares with robust standard errors. The number of observations is 163. YOUNG70 and YOUNG70 squared are always jointly significant at the 5-percent level. * and ** denote estimates that are significantly different from zero at the 10 and 5-percent level. The quadratic specification for YOUNG70 implies that the marginal effect of YOUNG70 on the increase in average schooling (average experience) would be negative (positive) for small values of YOUNG70; for sample values the effect is always positive (negative) however.

Table 2: Average-schooling externalities at the city-level: constant-composition approach 1970-1990

	All workers									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Constant	-0.12** (0.05)	-0.083** (0.02)	-0.21** (0.076)	-0.16** (0.05)	-0.16** (0.05)	-0.15** (0.05)	-0.16** (0.06)	-0.16** (0.03)	-0.16** (0.05)	-0.16** (0.04)
Change in average schooling 1970-1990 (ΔS)	0.014 (0.03)	-0.004 (0.017)	0.005 (0.034)	-0.01 (0.018)	0.003 (0.029)	-0.001 (0.018)	-0.006 (0.032)	-0.005 (0.028)	-0.006 (0.018)	0.001 (0.033)
Change in average experience 1970-1990 (ΔE)			-0.018** (0.008)	-0.017** (0.008)	-0.013 (0.009)	-0.017** (0.008)	-0.014 (0.010)	-0.014 (0.010)	-0.012 (0.017)	-0.015 (0.010)
Log-change in aggregate employment ($\Delta \log L$)	0.16** (0.06)	0.081** (0.027)	0.2** (0.07)	0.11** (0.04)	0.11** (0.05)	0.11** (0.05)	0.12** (0.04)	0.12** (0.05)	0.12** (0.04)	0.11** (0.04)
MOUNTAIN	-0.11** (0.02)	-0.09** (0.01)	-0.11** (0.03)	-0.09** (0.008)	-0.09** (0.008)	-0.09** (0.008)	-0.09** (0.009)	-0.09** (0.008)	-0.09** (0.01)	-0.09** (0.008)
MIDWEST	-0.07** (0.013)	-0.06** (0.01)	-0.06** (0.015)	-0.05** (0.011)	-0.05** (0.01)	-0.05** (0.011)	-0.053** (0.011)	-0.053** (0.011)	-0.051** (0.011)	-0.05** (0.011)
SOUTH	-0.03 (0.02)		-0.036 (0.026)							
WEST	-0.027 (0.03)		-0.032 (0.033)							
YOUNG70*AA70					-0.06 (0.1)					-0.35 (0.4)
YOUNG70						-0.022 (0.036)				
AA70							-0.04 (0.09)			0.28 (0.33)
AGE50P*AA70								-0.004 (0.34)		
AGE50P									0.13 (0.37)	
P-Value overidentifying restrictions			0.53	0.41						

Notes: 2SLS estimation with robust standard errors of (24) at the city-level. Constant-composition average wages constructed with data on all workers. Right-hand-side variables used are those in the leftmost column. Instruments used are: people in the city younger than 18 per adult in 1970 (YOUNG70), the share of the city-workforce older than 50 in 1970 (50PLUS70), the share of African-Americans in the city-population in 1970 (AA70), YOUNG70*YOUNG70, YOUNG70*AA70, YOUNG70*50PLUS70, and four macro-region dummies. * and ** denote estimates that are significantly different from zero at the 10 and 5-percent level

Table 3: Average-schooling externalities at the city-level: constant-composition approach 1970-1990

	White males					White males aged 40-49	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	-0.25** (0.1)	-0.17** (0.07)	-0.18** (0.07)	-0.18** (0.07)	-0.17** (0.08)	-0.27** (0.11)	-0.25** (0.068)
Change in average schooling 1970-1990 (ΔS)	0.046 (0.04)	-0.001 (0.021)	-0.017 (0.045)	-0.005 (0.041)	-0.001 (0.023)	0.07 (0.06)	-0.015 (0.03)
Change in average experience 1970-1990 (ΔE)	-0.02* (0.011)	-0.02* (0.011)	-0.02* (0.013)	-0.021* (0.013)	-0.02* (0.011)	-0.013 (0.015)	-0.027** (0.012)
Log-change in aggregate employment ($\Delta \log L$)	0.16** (0.08)	0.13** (0.05)	0.14** (0.05)	0.14** (0.05)	0.14** (0.05)	0.16* (0.1)	0.14** (0.05)
MOUNTAIN	-0.11** (0.03)	-0.12** (0.013)	-0.12** (0.012)	-0.12** (0.013)	-0.12** (0.014)	-0.035 (0.059)	
MIDWEST	-0.029 (0.019)	-0.035** (0.014)	-0.03** (0.01)	-0.035** (0.014)	-0.035** (0.014)	-0.029 (0.028)	
SOUTH	-0.018 (0.028)					-0.062 (0.047)	
WEST	0.026 (0.052)					0.018 (0.059)	
YOUNG70*AA70							
YOUNG70					-0.005 (0.05)		
AA70			-0.055 (0.14)				
AGE50P*AA70							
AGE50P				0.017 (0.14)			
P-Value overidentifying restrictions	0.87	0.73				0.31	0.22

Notes: 2SLS estimation with robust standard errors of (24) at the city-level. Constant-composition average wages constructed with data on white males only in columns (1) to (5) and white males aged 40-49 in columns (7) and (8). Right-hand-side variables used are those in the leftmost column. Instruments used are: people in the city younger than 18 per adult in 1970 (YOUNG70), the share of the city-workforce older than 50 in 1970 (50PLUS70), the share of African-Americans in the city-population in 1970 (AA70), YOUNG70*YOUNG70, YOUNG70*AA70, YOUNG70*50PLUS70, and four macro-region dummies. The number of observations is 163. * and ** denote estimates that are significantly different from zero at the 10 and 5-percent level.

Table 4: Average-schooling externalities: constant-composition approach 1970-1990 with restricted scale effects

	(1)	(2)	(3)
	\hat{a}	\hat{a}	\hat{a}
$d = 0.04$	-0.008 (0.026)	-0.017 (0.027)	-0.022 (0.016)
$d = 0.06$	-0.004 (0.025)	-0.014 (0.027)	-0.019 (0.016)
$d = 0.08$	0.0002 (0.025)	-0.01 (0.028)	-0.015 (0.016)
$d = 0.1$	0.0004 (0.026)	-0.006 (0.028)	-0.011 (0.016)

Notes: 2SLS estimation with robust standard errors of the strength of average-schooling externalities at the city-level using (24) restricting the strength of aggregate scale effects d to the values in the leftmost column. Instruments used are: people in the city younger than 18 per adult in 1970 (YOUNG70), the share of the city-workforce older than 50 in 1970 (50PLUS70), the share of African-Americans in the city-population in 1970 (AA70), YOUNG70*YOUNG70, YOUNG70*AA70, YOUNG70*50PLUS70, and four macro-region dummies. The number of observations is 163. * and ** denote estimates that are significantly different from zero at the 10 and 5-percent level. The control variables used are:

- Column (1): a constant and four macro-region dummies. The P-value of the hypothesis that the macro-region dummies SOUTH and WEST can be excluded from the estimating equation is 0.43.
- Column (2): same as in (1) plus the increase in average experience. The P-value of the hypothesis that the macro-region dummies SOUTH and WEST can be excluded from the estimating equation is 0.57.
- Column (3): a constant and two macro-region dummies (MOUNTAIN, MIDWEST) plus increase in average experience. The P-values of the test of overidentifying restrictions (not in the table) indicate that these restrictions cannot be rejected at standard significance levels for the values of d in the table.

Table 5: Average-schooling externalities at the state-level: constant-composition approach 1970-1990

	White males aged 40-49	
Change in average schooling at the US state-level 1970-1990 (ΔS)	0.029 (0.028)	0.027 (0.026)
P-Value overidentifying restrictions	0.86	0.66
Comments	AAIV	OURIV

Notes: 2SLS estimation at the US state-level of $\Delta \log \hat{w}_s^F = \text{Constant} + a\Delta S_s$ for 1970-1990, where state-level constant-composition average wages are constructed using white males aged 40-49 only. AAIV refers to the instruments used by Acemoglu and Angrist (2000) and OURIV refers to our instruments used at the state-level. The number of observations is 49. See the main text for additional explanations.

Table 6: Effect of change in average schooling 1970-1990 on mover-stayer wage-differential (\hat{f}_{cst})

	Years of schooling of workers (s)					
	[0,9)	[9,12)	[12,14)	[14,16)	[16,17)	[17 and more)
Sample A, 1980, least squares (LS)	-0.024 (0.058)	0.034* (0.022)	0.037* (0.015)	0.092* (0.026)	0.038** (0.019)	-0.014 (0.031)
Sample A, 1980, 2SLS	0.13 (0.15)	0.023 (0.035)	0.005 (0.024)	0.075 (0.04)	0.037 (0.028)	0.09 (0.13)
Sample B, 1980, LS	-0.04 (0.064)	0.038* (0.024)	0.04** (0.016)	0.09** (0.03)	0.035 (0.021)	-0.027 (0.038)
Sample B, 1980, 2SLS	0.11 (0.1)	0.027 (0.038)	0.01 (0.026)	0.068 (0.046)	0.04 (0.032)	0.045 (0.15)
Sample A, 1990, LS	-0.12 (0.09)	0.057* (0.036)	0.057** (0.018)	0.014 (0.029)	0.016 (0.03)	0.012 (0.037)
Sample A, 1990, 2SLS	-0.16 (0.13)	0.028 (0.056)	0.062** (0.029)	0.036 (0.026)	0.047 (0.047)	0.032 (0.07)
Sample B, 1990, LS	-0.12 (0.1)	0.077* (0.043)	0.066** (0.021)	0.025 (0.036)	0.063 (0.051)	0.1* (0.056)
Sample B, 1990, 2SLS	-0.13 (0.13)	0.098 (0.068)	0.062* (0.033)	0.056 (0.056)	0.048 (0.063)	0.1 (0.08)

Notes: 2SLS estimation with robust standard errors of the effect of the change in average schooling 1970-1990 on the mover-stayer wage-differentials \hat{f}_{cst} , $t = 1980, 1990$, at the city-level using $\hat{f}_{cst} = \text{Controls} + d\Delta \log L_{c,70-90} + a\Delta S_{c,70-90}$. The mover-stayer wage-differentials \hat{f}_{cst} are estimated using (26) for the six individual schooling intervals heading the table. Control variables used are a constant and four (of the five) macro-region dummies as well as the change in average experience 1970-1990. Sample A refers to all workers who did not live in the city five years before we observe their wages (“movers”) plus workers who lived in the same house since 1970. Sample B refers all workers who did not live in the city five years before we observe their wages (“movers”) plus workers who already lived in the same city five years before and were born in the state of residence. Instruments used for 2SLS estimation are: people in the city younger than 18 per adult in 1970 (YOUNG70), the share of the city-workforce older than 50 in 1970 (50PLUS70), the share of African-Americans in the city-population in 1970 (AA70), YOUNG70*YOUNG70, YOUNG70*AA70, YOUNG70*50PLUS70, and four macro-region dummies. The number of observations is 163. * and ** denote estimates that are significantly different from zero at the 10 and 5-percent level.

Table 7: Average-schooling externalities: Mincerian approach 1970-1990

	All workers			White males
	(1)	(2)	(3)	(4)
Constant	-0.96** (0.08)	-0.88** (0.05)	-1.14** (0.17)	-0.62** (0.09)
Change in average schooling 1970-1990 (ΔS)	0.085** (0.036)	0.071** (0.017)	0.23** (0.06)	0.079** (0.038)
Change in average experience 1970-1990 (ΔE)	-0.017* (0.01)	-0.013* (0.008)	-0.012 (0.025)	-0.016* (0.01)
Log-change in aggregate employment ($\Delta \log L$)	0.28** (0.08)	0.16** (0.04)	0.36** (0.13)	0.22** (0.09)
MOUNTAIN	-0.09** (0.026)	-0.053** (0.013)	-0.084** (0.041)	-0.16** (0.03)
MIDWEST	0.066** (0.016)	-0.055** (0.012)	-0.069** (0.034)	-0.054** (0.017)
SOUTH	-0.047 (0.035)			
WEST	-0.047 (0.036)			
P-Value overidentifying restrictions	0.64	0.21	0.27	0.23
Individual return assumed identical across cities?	Yes	Yes	No	Yes

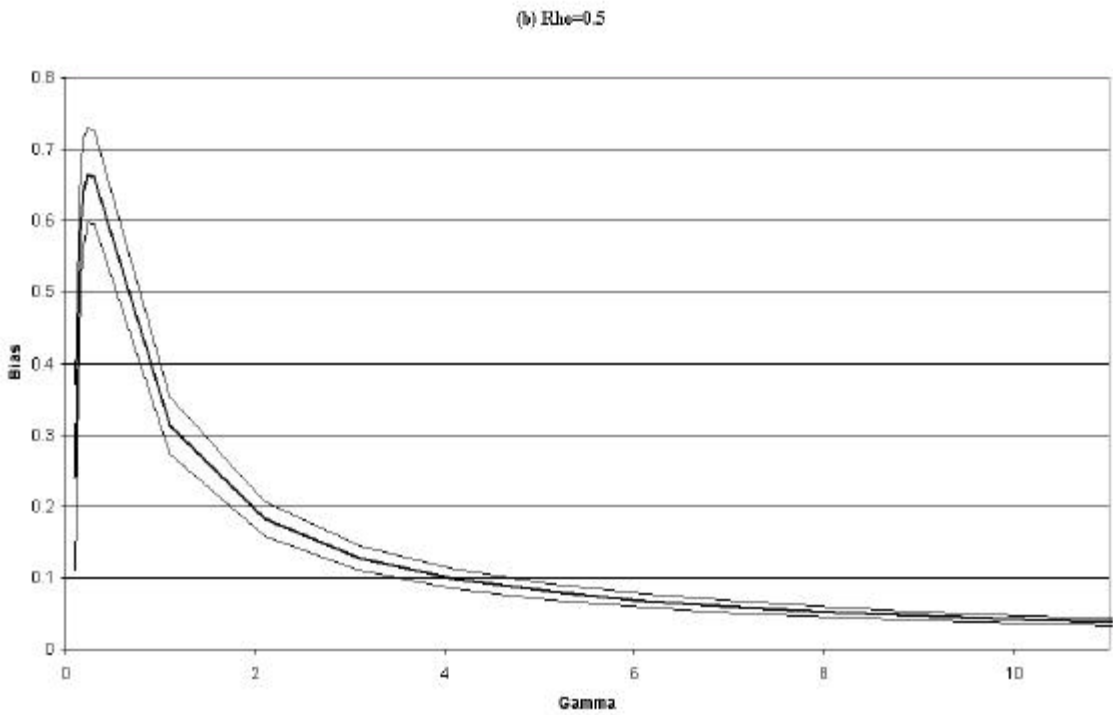
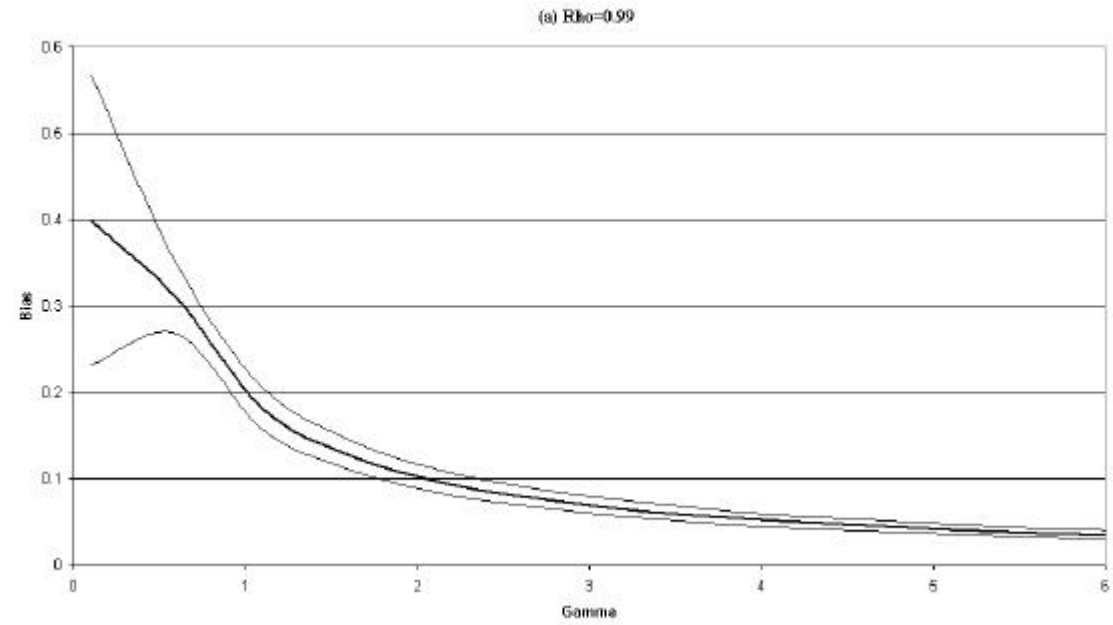
Notes: 2SLS estimation with robust standard errors of (20) at the city-level. Mincerian intercepts are estimated with data on all workers in columns (1) to (3) and white males only in column (4). Right-hand-side variables used are those in the leftmost column. Instruments used are: people in the city younger than 18 per adult in 1970 (YOUNG70), the share of the city-workforce older than 50 in 1970 (50PLUS70), the share of African-Americans in the city-population in 1970 (AA70), YOUNG70*YOUNG70, YOUNG70*AA70, YOUNG70*50PLUS70, and four macro-region dummies. The number of observations is 163. * and ** denote estimates that are significantly different from zero at the 10 and 5-percent level.

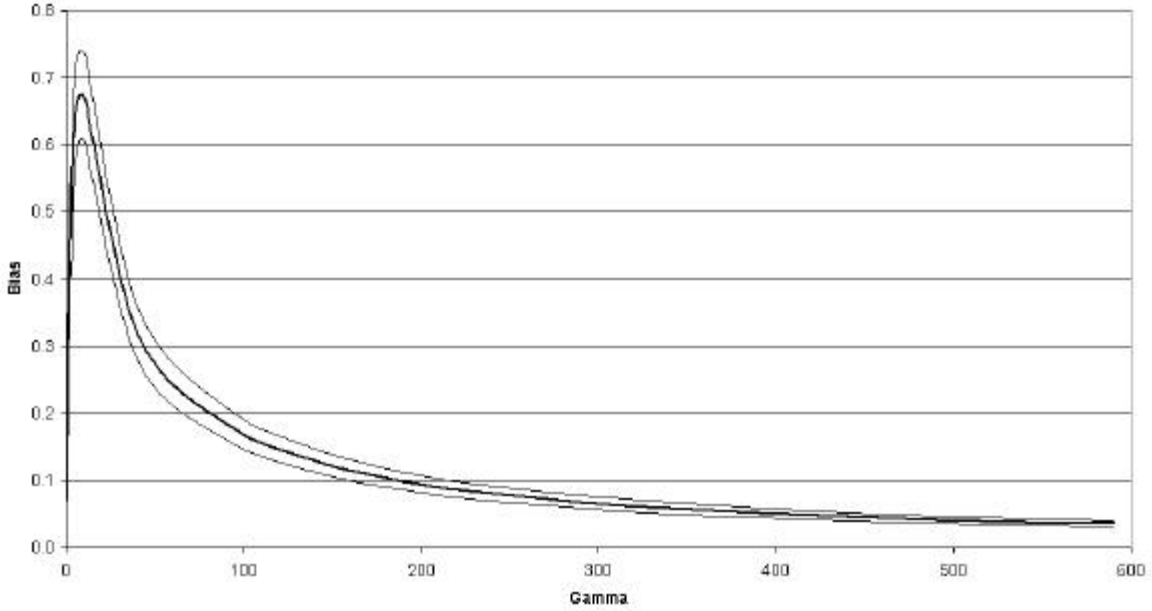
Table 8: Effect of average schooling on the individual return to schooling 1970-1990

	All workers											White males
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Constant	0.03** (0.01)	0.05** (0.005)	0.06** (0.01)	0.055** (0.015)	0.035** (0.007)	0.045** (0.006)	0.05** (0.008)	0.055** (0.015)	0.059** (0.017)	0.06** (0.01)	0.055** (0.011)	0.052** (0.01)
Change in average schooling 1970-1990 (ΔS)	-0.006 (0.007)	-0.015** (0.004)	-0.021** (0.007)	-0.015** (0.004)	-0.015** (0.004)	-0.014** (0.003)	-0.011** (0.004)	-0.01** (0.004)	-0.01** (0.004)	-0.013** (0.006)	-0.013* (0.007)	-0.013** (0.006)
Change in average experience 1970-1990 (ΔE)	-0.0011 (0.001)	-0.001 (0.001)	-0.002 (0.002)	-0.001 (0.0012)	0.004 (0.003)							
Log-change in aggregate employment ($\Delta \log L$)	-0.035** (0.01)	-0.021** (0.005)	-0.022** (0.006)	-0.021** (0.004)	-0.021** (0.006)	-0.022** (0.006)	-0.039** (0.01)	-0.041** (0.011)	-0.041** (0.015)	-0.04** (0.01)	-0.04** (0.011)	-0.037** (0.012)
Macro-region controls?	Yes	No	No	No	No	No	No	No	No	No	No	No
YOUNG70*AA70											0.006 (0.03)	
YOUNG70				-0.002 (0.008)				-0.01 (0.011)				
AA70			0.027 (0.026)							-0.007 (0.02)		
AGE50P*AA70												
AGE50P					0.12 (0.09)				-0.021 (0.06)			
P-Value overidentifying restrictions	0.74	0.83				0.61	0.81					0.75
Comments							Macro-region dummies not used as instruments					

Notes: 2SLS estimation with robust standard errors of (21) at the city-level. Individual returns to schooling are estimated with data on all workers in columns (1) to (11) and white males only in column (12). Right-hand-side variables used are those in the leftmost column. Instruments used are: people in the city younger than 18 per adult in 1970 (YOUNG70), the share of the city-workforce older than 50 in 1970 (50PLUS70), the share of African-Americans in the city-population in 1970 (AA70), YOUNG70*YOUNG70, YOUNG70*AA70, YOUNG70*50PLUS70, and four macro-region dummies. The number of observations is 163.* and ** denote estimates that are significantly different from zero at the 10 and 5-percent level.

Figure 1: Signing the bias if second-order effects are important



(c) $\text{Rho}=0.01$ 

Notes: Figures (a) to (c) display the estimates of the coefficient on ΔS_c (and the 95-percent confidence bands) when estimating $\mathbf{b}_c(1 - \mathbf{b}_c)(\Delta S_c)^2 = \text{Controls} + \mathbf{d}\Delta \log L_{c,70-90} + \mathbf{a}\Delta S_{c,70-90}$ at the city-level. Control variables used are a constant and four (of the five) macro-region dummies as well as the change in average experience 1970-1990. The method of estimation is 2SLS with robust standard errors and the usual instruments. The values of $\mathbf{b}_c(1 - \mathbf{b}_c)$ underlying the figures are calculated using (25) assuming that \mathbf{r} is the value heading the figures and that \mathbf{g} satisfies, $\max_c \hat{b}_c \leq \mathbf{g} \leq [(100/\mathbf{r}) - (1 - \mathbf{r})/\mathbf{r}]$ where \hat{b}_c is the 1970 return to schooling in cities. For additional explanations see the main text.

A.2 Data and Some Statistics

The data used in the empirical analysis comes from the “Census of Population and Housing” PUMS files. For 1970 we have used the 5-percent sample modifying the extraction code kindly provided by David Card. The geographic identifier used for 1970 is the “County Group Code”. For 1980 and 1990 we have used the “Card and Chay” extracts of the 5-percent PUMS (available at <ftp://elsa.berkeley.edu/pub/census/>), which include the standard metropolitan statistical area (SMSA) code as a geographic identifier.

Construction of Cities

The definition of cities that we use corresponds, with some exceptions, to the US Bureau of Census definition of standard metropolitan statistical areas (SMSAs) in 1990. The PUMS of the 1980 and 1990 US Census have (FIPS) codes identifying the SMSA where individuals live. With this information we can assign individuals in 1980 and 1990 to one of 236 cities. The 1970 US Census does not identify the SMSAs where individuals live, only whether they live in a SMSA or not. Individuals are instead assigned to so-called county groups. County groups can be related to SMSAs by using the so-called county group map (attached to the PUMS in 1970). We match individuals to SMSAs in the following way. When one or more county groups were contained in one SMSA, then we assign individuals located in one of the county groups to the SMSA that contains them. When a county group contained more than one SMSA, then we merged the different SMSAs into one (13 of our 163 cities are obtained this way) applying the same

criterion to SMSAs in 1980 and 1990 (to ensure that cities are defined in the same way in 1970, 1980, and 1990). Finally, when a county group was contained partly in a SMSA and partly in a non-SMSA area, then we assigned all individuals located in the county group who lived in a SMSA to the SMSA that contained part of the county group. This procedure resulted in 163 cities for 1970, 1980, and 1990. The code to perform the identification and merge of cities is available from us upon request.

Definition of Individual Wages and Schooling

Hourly wages in a given year have been calculated as yearly salary and wage divided by weeks worked times average hours per week worked in the year. All regressions are run using only individuals with positive hours worked and non-negative potential experience. All variables refer to the year previous to the census. Top-codes differ across years. Individual years of schooling have been obtained in the following way. For 1970 and 1980, we use the variables “Highest Grade Attended” and “Grade” which yields nineteen levels of schooling. For 1990, we use the variable “Yearsch” which yields eleven levels of schooling. When we only observe an interval for years of schooling, we use the midpoint of the interval as years of schooling in the Mincerian wage regressions in (19). Years of experience used in (19) is potential experience, i.e. age minus years of schooling minus six. For the constant-composition approach in (23) we partition years of schooling in seven intervals [0-4), [4-9), [9-12), [12-14), [14-16), [16-17), and [17 and more) and years of experience in five intervals [0-10), [10-20), [20-30), [30-40), [40-more).

Wages and Workforce Composition by Schooling

Sample values for average hourly wages in 1990 US\$ and workforce shares for the 7 education groups (at the national level) are given in the next table.

Years of Schooling	1970		1980		1990	
	Share	Average Wage	Share	Average Wage	Share	Average Wage
[0-4)	0.026	10.9	0.016	10.6	0.013	9.8
[4-9)	0.11	11.1	0.05	11.1	0.03	10.5
[9-12)	0.18	11.5	0.12	10.6	0.12	10.3
[12-14) (High School)	0.43	12.1	0.45	10.8	0.51	10.9
[14-16)	0.096	14.3	0.14	12.1	0.074	12.8
[16-17) (College)	0.08	18.7	0.11	14.7	0.17	16.4
[17 and more)	0.07	20.7	0.10	16.9	0.085	20.6

A.3 Cities and Key Aggregate Variables

City Name	Average Years of Schooling		Average Years of Experience		Total Private, Non-Agricultural Employment	
	'70	'90	'70	'90	'70	'90
Abilene, TX	11.6	13.2	24.1	18.4	54707	91746
Akron, OH	12.1	13.1	22.7	18.8	237978	285068
Albany-Schenectady-Troy, NY	12.3	13.4	23.6	19.5	262619	384949
Albuquerque, NM	12.7	13.3	21.0	18.7	100206	244273
Allentown-Bethlehem-Easton, PA-NJ	11.3	12.9	25.3	20.0	238793	303955

Altoona, PA	11.4	12.6	24.1	19.5	49094	55627
Amarillo, TX	12.0	12.9	21.2	19.3	53301	82711
Appleton-Oshkosh-Neenah, WI	11.7	12.9	22.3	18.3	99351	163027
Atlanta, GA	11.9	13.5	20.6	17.9	703363	1580647
Atlantic City, NJ	10.8	12.6	25.7	20.4	131011	236189
Augusta, GA-SC	11.1	12.9	22.5	19.1	88932	177811
Austin, TX	12.4	13.5	20.5	17.1	108915	351303
Bakersfield, CA	11.6	12.9	23.6	19.8	84895	187927
Baltimore, MD	11.2	13.2	23.4	18.7	748907	1136605
Baton Rouge, LA	12.2	12.9	21.7	17.4	108653	217277
Beaumont-Port Arthur, TX	11.2	12.7	23.3	20.0	116527	143500
Billings, MT	12.2	13.5	22.6	17.7	54915	91819
Biloxi-Gulfport, MS	11.5	12.8	22.6	19.9	39027	68701
Binghamton, NY	12.3	13.4	22.7	18.8	92062	117649
Birmingham, AL	11.4	13.0	23.0	19.6	283762	418355
Bloomington-Normal, IL	12.3	13.5	22.3	17.9	36163	66448
Boise City, ID	12.8	13.3	20.6	17.6	41377	112395
Boston, MA	12.3	13.8	23.8	18.1	1716968	2314349
Bridgeport, CT	11.6	12.9	25.2	21.1	340672	447286
Brownsville-Harlingen, TX	9.2	10.7	24.9	20.3	70841	179994
Buffalo, NY	11.7	13.2	24.4	19.8	385600	450640
Canton, OY	11.6	12.6	22.9	19.1	144914	176241
Cedar Rapids, IA	12.3	13.2	20.7	19.8	68057	98237
Champaign-Urbana-Rantoul, IL	13.3	13.9	18.5	17.9	38843	72901
Charleston, SC	10.9	12.9	22.3	18.2	85389	195445
Charlotte-Gastonia-Rock Hill, NC-SC	11.8	12.7	21.4	18.5	380775	669127
Chattanooga, TN-GA	11.1	12.4	23.5	21.0	141885	202487
Chicago, IL	11.9	13.2	23.3	19.3	2620409	3262950
Cincinnati, OH-KY-IN	11.7	13.1	22.8	18.8	519052	768186
Cleveland, OH	11.9	13.1	23.6	19.5	848369	962555
Colorado Springs, CO	12.3	13.4	22.3	19.7	87520	207905
Columbia, MO	12.8	13.9	19.2	13.2	22604	50780
Columbia, SC	11.7	13.2	20.5	18.6	106882	224986
Columbus, OH	12.3	13.3	21.2	18.0	408761	712922
Corpus Christi, TX	10.8	12.6	23.5	18.4	86664	132307
Dallas, TX	11.8	13.1	21.4	17.9	987312	2082026
Davenport-Rock Island-Moline, IA-IL	11.8	13.1	23.6	18.4	128309	164751
Dayton-Springfield, OH	11.8	13.2	22.4	19.5	350730	442932
Decatur, IL	11.8	13.2	24.6	21.0	116410	154621
Denver, CO	12.6	13.4	20.9	18.8	437946	912957
Des Moines, IA	12.5	13.2	22.3	18.3	142745	242171
Detroit, MI	11.9	13.1	22.5	19.2	1578307	1993757
Duluth, MN-WI	12.1	12.8	24.5	18.8	81523	94318
El Paso, TX	11.1	11.9	22.2	19.0	99848	199834
Erie, PA	12.0	13.0	24.2	20.0	100555	126113
Eugene-Springfield, OR	12.6	13.2	21.5	19.3	65898	123967

Fayetteville, NC	11.1	13.0	20.5	16.3	42539	77154
Flint, MI	11.6	12.8	21.1	19.6	144957	171398
Fort Lauderdale-Hollywood-P.Beach, FL	11.7	13.0	24.4	20.3	206800	583406
Fort Wayne, IN	12.2	12.8	21.4	19.7	141088	212278
Fresno, CA	11.9	12.7	21.8	19.2	122350	257420
Gainesville, FL	12.9	13.9	19.8	16.2	29092	78446
Gary-Hammond, IN	11.4	12.8	23.6	20.5	227318	247774
Grand Rapids, MI	11.8	13.1	22.9	17.7	199974	396412
Green Bay, WI	11.9	12.8	21.3	18.2	55787	108272
Greensboro-Winston-Salem-High Point, NC	11.1	12.7	22.8	19.9	330504	537601
Greenville-Spartanburg, SC	10.9	12.7	22.2	19.9	202077	352503
Hamilton-Middletown, OH	11.3	12.8	22.7	18.2	68885	96991
Harrisburg-Lebanon-Carlisle, PA	11.6	12.8	23.7	19.2	191536	289589
Honolulu, HI	12.2	13.3	21.3	19.6	218323	388070
Houston, TX	11.8	12.8	20.8	18.2	829788	1728781
Huntington-Ashland, WV-KY-OH	11.7	12.8	24.4	20.6	93553	112138
Indianapolis, IN	11.9	12.9	22.3	18.6	428858	686737
Jackson, MI	11.8	12.6	23.6	19.7	48035	54326
Jackson, MS	12.1	13.8	21.4	18.5	106672	175415
Jacksonville, FL	11.6	12.7	22.7	19.3	211580	429792
Jersey City, NJ	10.4	12.2	26.3	20.8	252752	234772
Johnstown, PA	11.3	12.6	24.8	21.5	76736	85583
Kalamazoo, MI	12.4	13.4	20.9	18.3	70735	114059
Kansas City, MO-KS	12.1	13.2	22.8	19.1	525933	808325
Kenosha, WI	11.6	12.9	23.5	19.5	90329	121051
Knoxville, TN	11.6	12.6	23.4	21.0	149498	280887
Lafayette, LA	11.4	12.8	22.1	17.7	42068	103342
Lafayette, IN	12.6	13.5	22.7	14.7	36562	59149
Lancaster, PA	11.0	12.4	24.4	20.5	130428	215469
Lansing-East Lansing, MI	12.5	13.5	20.2	18.0	107632	177065
Las Vegas, NY	12.1	12.4	21.2	20.0	105346	390749
Lexington-Fayette, KY	12.4	13.5	19.5	17.7	96774	189803
Lima, OH	11.7	12.3	23.4	21.9	59761	76125
Lincoln, NE	12.8	13.5	21.0	16.9	62273	111918
Lorain-Elyria, OH	11.2	12.5	23.7	20.6	79172	93067
Los Angeles-Long Beach, CA	12.3	12.4	22.1	19.1	2867092	4617235
Louisville, KY-IN	11.4	13.2	23.2	20.2	356409	499585
Lubbock, TX	11.5	13.0	21.4	17.1	58604	97582
Macon-Warner Robins, GA	11.1	12.8	23.7	17.1	71820	115158
Madison, WI	13.0	14.1	20.0	16.9	94819	192086
Mansfield, OH	11.4	12.6	22.7	19.6	52239	62090
Memphis, TN-AR-MS	11.5	13.2	22.5	18.2	288933	490942
Miami-Hialeah, FL	11.5	12.5	24.2	21.4	552362	961182
Milwaukee, WI	12.0	13.2	22.7	19.0	569223	786156
Minneapolis-St. Paul, MN-WI	12.4	13.5	21.2	18.0	798927	1410586
Modesto, CA	12.0	12.4	22.0	18.5	57558	133340

Monroe, LA	11.5	12.9	23.6	17.8	35308	56537
Montgomery, AL	11.3	13.1	23.7	19.3	72900	121578
Muncie, IN	11.6	12.6	22.4	19.7	94302	103926
Nashville, TN	11.7	13.0	22.2	18.8	270000	537601
New Orleans, LA	11.4	13.3	23.2	19.4	400789	545158
New York, NY	11.8	13.1	24.3	20.5	4008941	4057008
Newark, NJ	11.7	13.3	24.5	21.3	814079	948960
Norfolk-Virginia Beach-Newport News, VA	11.3	13.0	22.7	18.8	277854	552493
Odessa, TX	11.8	12.0	21.9	18.7	59032	101496
Oklahoma City, OK	12.3	13.2	21.5	18.4	245768	423628
Omaha, NE-IA	12.2	13.3	21.7	18.8	214619	340565
Orlando, FL	11.7	13.1	22.5	18.7	154095	598302
Pensacola, FL	11.5	12.8	22.9	20.1	175921	307537
Peoria, IL	11.8	13.0	23.3	19.3	131764	159114
Philadelphia, PA-NJ	11.7	13.3	24.0	19.9	1782621	2268610
Phoenix, AZ	12.3	13.2	21.3	18.5	339996	1048770
Pittsburgh, PA	11.8	13.3	24.6	19.9	835990	956984
Portland, OR	12.5	13.4	22.2	18.5	359393	683078
Providence, RI	11.2	13.1	24.8	18.4	14104	17932
Raleigh-Durham, NC	11.9	14.0	20.8	17.5	166559	410551
Reading, PA	11.0	12.6	24.7	20.0	125883	162847
Reno, NY	12.5	13.1	21.9	19.5	56132	151804
Richmond-Petersburg, VA	11.4	13.2	22.9	18.7	265752	461658
Riverside-San Bernardino, CA	12.0	12.5	22.7	18.4	282666	774484
Roanoke, VA	11.6	12.6	23.0	21.0	86656	135770
Rochester, NY	12.1	13.5	22.7	18.8	355418	500441
Rockford, IL	11.4	12.7	23.3	19.9	111209	150339
Sacramento, CA	12.6	13.4	22.2	18.5	211398	594227
Saginaw-Bay City-Midland, MI	11.6	13.0	22.4	20.5	127609	167441
St. Louis, MO-IL	11.6	13.1	23.8	19.3	903707	1249521
Salem, OR	12.4	12.8	22.8	19.1	48702	101563
Salinas-Seaside-Monterey, CA	12.0	12.7	22.1	20.2	61677	139790
Salt Lake City-Ogden, UT	12.8	13.4	21.3	17.5	253032	575744
San Antonio, TX	11.1	12.8	22.6	18.6	244385	498204
San Diego, CA	12.7	13.3	21.7	17.2	363752	1061203
San Francisco, CA	12.7	13.8	22.2	19.0	1191491	2030555
San Jose, CA	12.9	13.6	19.7	17.7	373632	910406
Santa Barbara-Santa Maria-Lompoc, CA	12.8	13.1	21.5	19.0	78815	172008
Santa Rosa-Petaluma, CA	12.5	13.4	22.4	19.1	51310	169499
Seattle, WA	12.7	13.6	21.2	18.5	509810	1169623
Shreveport, LA	11.5	13.0	23.9	20.1	96858	135989
South Bend-Mishawaka, IN	11.6	12.9	24.5	18.5	91007	123928
Spokane, WA	12.6	13.4	22.7	17.1	91290	159042
Springfield, MO	12.0	12.9	21.0	17.0	67234	135646
Stockton, CA	11.7	12.4	24.3	19.7	83178	165017
Syracuse, NY	12.2	13.4	23.3	17.9	212629	316047

Tacoma, WA	12.1	13.1	23.5	18.6	98195	196004
Tampa-St. Petersburg-Clearwater, FL	11.8	13.0	23.9	20.3	353339	964453
Terre Haute, IN	11.8	12.9	24.2	19.1	47586	56424
Toledo, OH	11.8	13.0	23.0	18.5	232895	291353
Trenton, NJ	11.7	13.6	24.0	20.2	120430	165416
Tucson, AZ	12.3	13.2	22.9	17.6	102945	256183
Tulsa, OK	12.0	13.2	21.8	18.7	199945	354355
Tuscaloosa, AL	12.0	13.0	21.1	16.3	31720	55412
Tyler, TX	11.5	12.9	25.1	20.6	38249	71843
Utica-Rome, NY	11.7	13.1	24.9	19.8	102463	118367
Vallejo-Fairfield-Napa, CA	12.2	13.1	23.8	19.8	49785	137181
Waco, TX	11.3	12.5	25.2	20.6	52641	80474
Washington, DC-MD-VA	12.8	14.0	20.5	18.0	930498	2046568
Waterloo-Cedar Falls, IA	12.2	13.2	22.4	18.2	56209	68731
West Palm Beach-Boca Raton-D. Beach, FL	11.6	12.9	24.7	21.1	129228	422008
Wichita, KS	12.3	13.1	22.6	19.7	156370	261496
Wilmington, DE-NJ-MD	12.0	13.4	22.2	19.7	190549	303117
Wilmington, NC	11.4	13.0	22.1	18.6	35038	64160
Worcester, MA	11.5	13.2	26.3	18.7	49470	84359
York, PA	11.0	12.5	24.8	19.2	137163	194569

A.4 Extending the Model: Physical Capital and Land

In the model in the main text we assumed for simplicity that production uses (different types of) labor only. We now put land and physical capital into the production function. The national market for physical capital will be taken to be perfectly competitive. This assumption seems reasonable for the US and is convenient as there is no data on the aggregate physical capital stocks at the local geographic level.

Suppose that the aggregate production function is

$$Y_{ct} = ((A_{ct} F(L_{ct}, H_{ct}))^{1-e} K_{ct}^e)^{1-r} N_c^r, \quad \mathbf{e}, \mathbf{r} \in (0,1) \quad (\text{A1})$$

where K_{ct} , N_c denote the amount of physical capital and the fixed amount of land employed in production (we are assuming the Cobb-Douglas production function for simplicity only). Assume that physical capital moves across cities to equalize its rate of return. Then city-year specific competitive equilibrium prices of labor and human capital are

$$w_{ct}^L = (1 - \mathbf{e})(1 - \mathbf{r}) \mathbf{f}(r_t) A_{ct} (N_c / (A_{ct} F(L_{ct}, H_{ct})))^{r/(1-e(1-r))} F_1(1, h_{ct}) \quad (\text{A2})$$

$$w_{ct}^H = (1 - \mathbf{e})(1 - \mathbf{r}) \mathbf{f}(r_t) A_{ct} (N_c / (A_{ct} F(L_{ct}, H_{ct})))^{r/(1-e(1-r))} F_2(1, h_{ct}), \quad (\text{A3})$$

where r_t is the national rate of return to physical capital and $\mathbf{f}(\bullet)$ some unimportant function. Average labor productivity and the equilibrium average wage satisfy

$$\begin{aligned} w_{ct} &= (1 - \mathbf{e})(1 - \mathbf{r})(Y_{ct} / L_{ct}) \\ &= (1 - \mathbf{e})(1 - \mathbf{r}) \mathbf{f}(r_t) A_{ct} (N_c / (A_{ct} F(L_{ct}, H_{ct})))^{r/(1-e(1-r))} F(1, h_{ct}) \end{aligned} \quad (\text{A4})$$

It can be shown that (A2) to (A4) combined with (6) imply that the elasticity of average wages with respect to average human capital holding workforce skill-composition weights constant is $(\mathbf{q}(1-\mathbf{e})(1-\mathbf{r})-\mathbf{rb})/(1-\mathbf{e}(1-\mathbf{r}))$, which is the strength of average human capital externalities $\mathbf{q}(1-\mathbf{e})(1-\mathbf{r})$ on TFP minus congestion effects \mathbf{rb} divided by one minus the share of physical capital in income $\mathbf{e}(1-\mathbf{r})$ (\mathbf{b} is the share of human capital in wages).

A.5 Extending the Model: Non-Tradable Goods

Suppose that the production function is $Y_{ct} = (A_{ct} F(L_{yct}, H_{yct}))^{1-r} N_{yct}^r$, $\mathbf{r} \in [0,1)$ for the tradable composite good and $X_{ct} = (A_{xct} F(L_{xct}, H_{xct}))^{1-r} N_{xct}^r$ for the non-tradable composite good; N_y , N_x denote the amount of land used in the production of the tradable and non-tradable good respectively. Hence, the production functions are identical except that TFP in the production of the non-tradable good may differ from TFP in the production of the tradable good. Suppose also that both goods are produced under perfect competition and that the tradable good is the numeraire. The equilibrium wage schedule in this case is $w_{ct}(H) = w_{ct}^L + w_{ct}^H H$ and prices of labor and human capital satisfy

$$w_{ct}^L = (1-\mathbf{r})A_{ct}^{1-r} F(1, h_{yct})^{-r} (N_{yct}/L_{yct})^r F_1(1, h_{yct}) \quad (\text{A5})$$

$$w_{ct}^H = (1-\mathbf{r})A_{ct}^{1-r} F(1, h_{yct})^{-r} (N_{yct}/L_{yct})^r F_2(1, h_{yct}) . \quad (\text{A6})$$

The fact that producers of the non-tradable and the tradable good in each city face the same factor prices implies $h_{yct} = h_{xct} = h_{ct}$ and $N_{yct}/L_{yct} = N_{xct}/L_{xct} = N_c/L_{ct}$ where N_c is the fixed amount of land available in each city. Hence, prices of labor and human capital in each city satisfy

$$w_{ct}^L = (1-\mathbf{r})A_{ct}^{1-r} F(1, h_{ct})^{-r} (N_c/L_{ct})^r F_1(1, h_{ct}) \quad (\text{A7})$$

$$w_{ct}^H = (1-\mathbf{r})A_{ct}^{1-r} F(1, h_{ct})^{-r} (N_c/L_{ct})^r F_2(1, h_{ct}) . \quad (\text{A8})$$

The approach in Appendix A4 can therefore be used to identify externalities net of congestion effects in the production of the tradable (but not the non-tradable) good.

A.6 Extending the Model: Non-Tradable Intermediate Inputs Produced with Increasing Returns to Scale

The model in the main text concentrates on technological externalities. We now develop a simple model with an endogenous variety of non-tradable intermediate inputs produced with increasing returns to scale and ‘‘pecuniary’’ instead of technological externalities. Suppose that perfectly competitive firms produce a tradable good (which will be the numeraire) according to

$$Y = G(S_y, F(L_y, H_y)) \quad (\text{A9})$$

where both $F(\bullet)$ and $G(\bullet)$ are subject to constant returns to scale and S is a non-tradable intermediate-input composite (time-subscripts are suppressed throughout and y -subscripts denote factors used in the production of the tradable good). The intermediate-input composite is produced according to

$$S = \left(\int_0^n s_i^{(s-1)/s} di \right)^{s/(s-1)}, \quad s > 1 \quad (\text{A10})$$

where s_i is the amount of non-tradable intermediate input i used and n the variety of intermediate inputs available in the city. Non-tradable intermediate inputs are produced in a monopolistically competitive sector according to

$$s_i = \max[G(S_i, F(L_i, H_i)) - \mathbf{k}, 0], \quad (\text{A11})$$

where S_i , L_i , H_i denote intermediate-input composites, labor, and human capital used in the production of intermediate input i ; \mathbf{k} is the overhead resource-requirement of production. Each intermediate input is produced by a different firm and there is free entry of firms into the intermediate-inputs sector. It can be shown that the equilibrium wage schedule of this model is

$$w(x) = w^L + w^H x \quad (\text{A12})$$

where

$$w^L = d(L, h)F_1(1, h), \quad (\text{A13})$$

$$w^H = d(L, h)F_2(1, h), \quad (\text{A14})$$

with L aggregate employment and h average human capital in the city. Furthermore, average wages are equal to

$$w = d(L, h)F(1, h). \quad (\text{A15})$$

The function $d(L, h)$ is increasing in both arguments and captures the effect of aggregate employment and average human capital in cities on the equilibrium variety of specialized inputs and hence aggregate productivity. It can be seen by comparing (A13) to (A15) with (4) to (6) that the constant-composition approach can be used to identify the elasticity of $d(L, h)$ with respect to aggregate employment and average human capital.

A.7 Effects of Labor Supply on Relative Wages

Notice that (3) to (5) can be written as $w(x) = AF(1, h) + AF_2(1, h)(x - h)$ using constant returns to scale given TFP of (1). Hence,

$$\frac{\partial \frac{w(x^l)}{w(x^h)}}{\partial h} = - \frac{F(1, h)F_{22}(1, h)(x^h - x^l)}{\left(F(1, h) + F_2(1, h)(x^h - h) \right)^2} \quad (\text{A16})$$

and

$$\left. \frac{\partial \frac{w(x^l)}{w(x^h)}}{\partial L(x^l)} \right|_{L(x^h)+L(x^l)=\text{constant}} = \frac{F(1, h)}{\left(F(1, h) + F_2(1, h)(x^h - h) \right)^2} F_{22}(1, h)(x^h - x^l)^2, \quad (\text{A17})$$

where $L(x)$ denotes the amount of labor with human capital x used in production. The increase in the relative wage of low human capital workers is therefore proportional to $-F_{22}(1, h)(x^h - x^l)^2$ when the supply of low human capital workers decreases and the supply of high human capital workers increases by the same amount.

A.8 The Constant-Composition Approach to Human Capital Externalities at the Country-Level

The constant-composition approach as developed so far cannot be applied to the identification of human capital externalities at the country-level because it would be unrealistic to assume that all countries have access to perfectly competitive international physical capital markets (if it was possible to make this assumption, the approach in Appendix A4 could be adapted to the country-level). This raises the question of how the strength of average human capital externalities can be estimated at the country-level. To answer this question suppose that the production function at the country-level is

$$Y = AF(L, H, K) \quad (\text{A18})$$

where K is the physical capital stock used in production and the level of TFP is $A = Bh^q$ where B captures exogenous differences in TFP and q the strength of average human capital externalities at the country-level; L, H are defined as usual. Assume that the aggregate production function is twice continuously differentiable and subject to constant returns to scale to L, H, K as well as constant or decreasing returns to human capital and to physical capital. Suppose also that labor markets and the market for physical capital at the country-level are perfectly competitive and that firms maximize profits taking the level of TFP as given. Denote the rental price of physical capital at the country-level by r and define “factor income per worker” by $\int w(x, h)l(x)dx + rk$, where k is the physical capital intensity. Then the following proposition holds.

Proposition A1: The elasticity of factor income per worker with respect to the average level of human capital yields the strength of average human capital externalities when workforce skill-composition weights $l(x)$ and the physical capital intensity k are held constant

$$\left. \frac{\partial \left(\int w(x, h)l(x)dx + rk \right)}{\partial h} \frac{h}{\int w(x, h)l(x)dx + rk} \right|_{l(x) \forall x \text{ and } k \text{ constant}} = q. \quad (\text{A19})$$

Proof: The argument is very similar to the proof of Proposition 3, which is why we will only sketch the main elements. Competitive factor markets at the country-level, profit-maximization, and the aggregate production function imply that factor income per worker can be written as $\int w(x, h)l(x)dx + rk = AF_1 + AF_2h + AF_3k$, where F_i denotes the partial derivative of $F(L, H, K)$ with respect to the i -th argument. Hence, (A19) follows if $\partial(F_1 + F_2h + F_3k)/\partial h = F_{12} + F_{22}h + F_{32}k = 0$. To demonstrate this last equality, notice that constant returns to scale to L, H, K and twice continuous differentiability of the production function imply $F_{21}(1, h, k) + F_{22}(1, h, k)h + F_{23}(1, h, k)k = F_{12}(1, h, k) + F_{22}(1, h, k)h + F_{32}(1, h, k)k = 0$. **Q.E.D.**