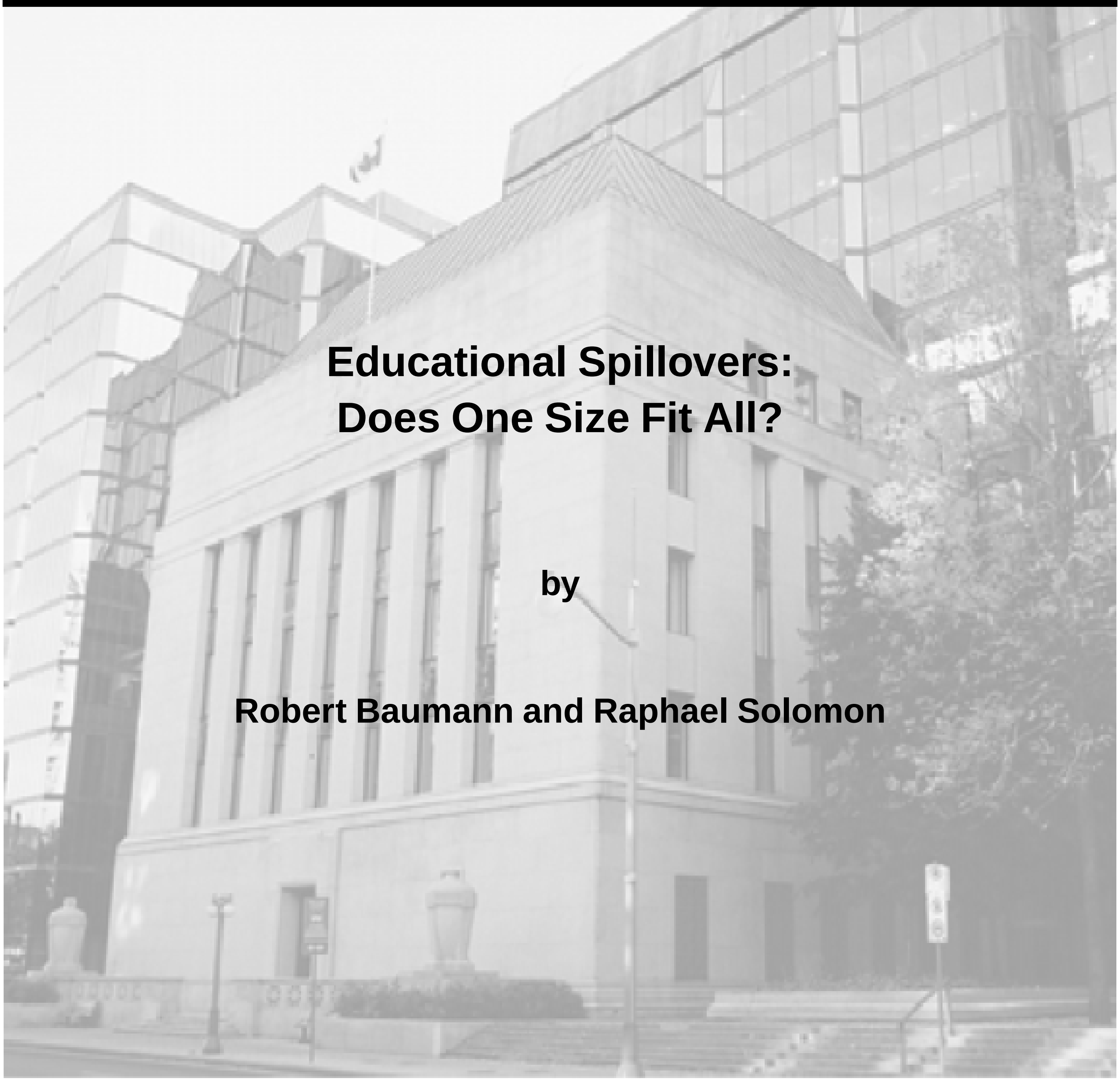
Bank of Canada Banque du Canada

Working Paper 2005-10 / Document de travail 2005-10



Educational Spillovers: Does One Size Fit All?

by

Robert Baumann and Raphael Solomon

ISSN 1192-5434

Printed in Canada on recycled paper

Bank of Canada Working Paper 2005-10

April 2005

Educational Spillovers: Does One Size Fit All?

by

Robert Baumann[1] and Raphael Solomon[2]

[1]Department of Economics
College of the Holy Cross
Worcester, MA 01610
rbaumann@holycross.edu

[2]Monetary and Financial Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
rsolomon@bankofcanada.ca

Contents

Acknowledgements

The authors thank Miles Cahill, John F. O’Connell, Francisco Covas, Patricia Reagan, Pierre St-Amant, Eric Santor, and Carolyn Wilkins for their helpful suggestions. The authors also appreciate comments from presentations at Clark University Economics Department and at the 31st Eastern Economic Association Annual Meetings. The responsibility for any errors remains with the authors.

Abstract

In a search model of production, where agents accumulate heterogeneous amounts of human capital, an individual worker's wage depends on average human capital in the searching population. Following this model, the authors use a large American panel data set to estimate a Mincerian wage equation augmented with terms for average human capital. They find that there is a positive and significant spillover effect, but that the effect differs by gender and population group (whites, blacks, and Hispanics), as well as educational status. The differing spillover effects can only partially be explained by occupational choice.

JEL classification: I29, J24, J31
Bank classification: Labour markets

Résumé

Dans un modèle de recherche d'emploi formalisant explicitement la production, où les agents accumulent des quantités hétérogènes de capital humain, le salaire individuel est fonction du capital humain moyen au sein de la population en recherche d'emploi. En utilisant ce modèle et un vaste ensemble de données de panel en provenance des États-Unis, les auteurs estiment une équation de salaire à la Mincer enrichie de termes représentant le capital humain moyen. Ils constatent l'existence d'un effet de débordement positif et significatif, dont l'ampleur varie selon le sexe, le groupe (Blancs, Noirs et personnes d'origine hispanique) et le niveau de scolarité atteint. Les variations de cet effet ne sont attribuables que partiellement au métier choisi.

Classification JEL : I29, J24, J31
Classification de la Banque : Marchés du travail

1. Introduction

It is widely accepted that wages are determined, at least in part, by the worker's human capital. Using education as a proxy for human capital, Mincer (1974) first formalized the relationship between education and wages at the individual level as an estimable equation, spawning a large literature. Human capital in the surrounding labour market affects individual wages, in addition to the effects of individual-level controls typically used in a Mincerian regression. While higher individual human capital raises wages, human capital in the surrounding labour market may increase or decrease individual wages. These labour market externalities interfere with the education-wage relationship at the individual level.

Labour market externalities exist because worker-firm bargains do not occur in a vacuum. As the supply of educated workers increases, holding all other factors constant, the "price" of educated workers, or their private return to education, drops. In the United States, the supply of educated workers has increased since at least 1940.[1] But all other factors are not necessarily equal. For example, two workers may produce more than twice the product of a single worker by sharing knowledge. Human capital accumulation in a given labour market expands the stock of information from which all workers can draw, raising individual worker productivity in the process. Presumably, the firm will prefer to keep some of this information private, but job mobility spreads information to rivals in the long run. Although it is not possible to identify the spillover and supply effects separately, a consistent estimate of the net effect of changes in a labour market's human capital has profound policy implications for public education. Public education subsidies exist (in part) because the social return to education is assumed to be larger than the private return (Friedman 1962). Without these subsidies, workers would acquire a suboptimally small stock of human capital.

In the extant empirical literature, there is no consensus on the semantics used to describe the positive effect of human capital accumulation on individual wages. Rudd (2000) and Moretti (2004a) call it a spillover, Acemoglu (1996) and Acemoglu and Angrist (2000) use the term social returns to education, Ciccone and Peri (2002) refer to the effect as an externality, and Rauch (1993) considers it a public good with a positive externality. At the heart of this debate is the propagation mechanism between human capital

[1] This draws on the authors' calculations from census data.

accumulation in a given labour market and individual wages. Our hypothesis is that worker interaction spreads information that affects productivity and wages. Marshall (1890), Lucas (1988), and Glaeser (1999), among others, discuss the effect on wages of human capital accumulation in the surrounding labour market. We feel the term "spillover" best describes this effect.

Three articles in the literature find evidence of positive spillovers. Rauch's (1993) was one of the first empirical studies of spillovers. Using 1980 U.S. Census Bureau data, Rauch finds positive and significant effects of city-level gains in mean human capital on wages. Moretti (2004a) uses data from the geocoded National Longitudinal Survey of Youth, 1979 (NLSY79), and defines labour markets using metropolitan statistical areas (MSAs). He treats labour market human capital as endogenous and finds positive and significant spillover effects. Moretti (2004b) extends his analysis by testing whether "economic distance" between industries affects spillovers. He uses three measures for economic distance: input/output tables, patent distribution across technological fields, and patent citations. Moretti (2004b) finds a positive spillover effect that varies by economic distance, and argues that these findings support the view that spillovers are related to worker interactions across industries.

Three other articles, however, cannot find evidence of spillover effects. Ciccone and Peri (2002) use U.S. Census Bureau data from 1970 and 1990 and also define a labour market as an individual MSA. They fit separate Mincerian regressions for each MSA and year. With these estimates, Ciccone and Peri regress the time-differenced constant term and the time-differenced coefficients for individual schooling on average schooling for each MSA. They conclude that there is no spillover effect. Rudd (2000) uses pooled cross-sections from 14 years of the March Current Population Survey and uses states as the local labour market. He includes a state-time fixed effect in a Mincerian model and regresses the fixed effects on various state characteristics, including average education. Rudd finds that average education is not a significant determinant of the fixed effect and concludes there are no spillover effects. Acemoglu and Angrist (2000) use data on males aged 40–49 from the 1960–80 census samples and states as the labour markets. They use instrumental variables for individual human capital and for average human capital, and find no evidence of a spillover effect.

We enter the ongoing debate and provide further evidence that the spillover effect dominates the supply effect. We extend the literature by estimating spillover effects across population groups (whites, blacks, and Hispanics), gender, and college completion status (graduate and non-graduate). To the best of our knowledge, we are the first to estimate separate spillover effects across population groups or gender.[2] We find that there are positive spillover effects and that these effects differ across groups. Our point estimates suggest that whites have the largest spillover effects, followed by Hispanics and blacks. Also, college graduates have larger spillovers than non-college graduates, and men have larger spillovers than women. Occupational choice accounts for roughly 20 per cent of the spillover differences. The remaining spillovers may be caused by differences in worker interactions across population groups, or gender.

The rest of this paper has the following structure. Section 2 sets out a theoretical framework within which to interpret our empirical results. The data are described in section 3. Section 4 presents our empirical model. The main results are described in section 5. Section 6 offers some conclusions.

2. A Theory of Spillovers

Our model combines aspects of Acemoglu and Angrist (2000) and Acemoglu (1996). A one-period economy consists of a large number of workers and a large number of firms. In the beginning, typical firm j invests in physical capital, k_j, at price r, while typical worker i invests in human capital, h_i. Then each firm meets one worker at random. After "bargaining," a share, β, of revenues accrues to workers and $(1-\beta)$ to firms. Finally, a random production shock, ε_j, occurs for each firm, with $\mathrm{E}\varepsilon_j = 1$. All firms have identical production possibilities, given by

$$y_{ij} = h_i^{\alpha} k_j^{1-\alpha} \varepsilon_j, \; 0 < \alpha < 1. \tag{1}$$

Workers have idiosyncratic preferences[3]:

$$u_i\left(h_i\right) = \beta h_i^{\alpha} k_j^{1-\alpha} - \frac{h_i^{1+\Gamma}}{g_i\left(1+\Gamma\right)}, \; \Gamma > 0. \tag{2}$$

The disutility of acquiring human capital is indexed both by a term common to all workers, Γ, and by an individual term, g_i, distributed uniformly

[2] Moretti (2004a) estimates separate spillover effects for college graduates and those without a college degree, and Moretti (2004b) estimates separate spillover effects across industries.

[3] This utility function is taken from Acemoglu (1996).

on [0,1].[4] Firms choose k_j to maximize expected profits[5] and workers choose h_i to maximize expected utility.[6] In the subgame perfect Nash equilibrium, all firms choose the same quantity of physical capital, namely

$$k^* = \left(\frac{r(\alpha + \eta)}{\psi c_0^{\alpha} \eta (1 - \beta)} \right)^{\frac{1}{\psi - 1}}, \tag{3}$$

where $\eta = \Gamma + 1 - \alpha > 0$, $c_0 = (\alpha\beta)^{\frac{1}{\eta}}$ and $\psi = \dfrac{(1-\alpha)(\alpha+\eta)}{\eta}$. Each worker chooses a different level of human capital:

$$h_i^* = g_i^{\frac{1}{\eta}} c_0 (k^*)^{\frac{1-\alpha}{\eta}}. \tag{4}$$

Workers' wages are given by:

$$w_{ij} = \beta (h_i^*)^{\alpha} (k^*)^{1-\alpha} \varepsilon_j. \tag{5}$$

Taking logs, we find that

$$\ln w_{ij} = \ln \beta + \alpha \ln h_i^* + (1 - \alpha) \ln k^* + \varepsilon_j. \tag{6}$$

Taking expectations over the distribution of g in (4), solving for k^* and substituting in (6) yields an augmented Mincerian regression:

$$\ln w_{ij} = c_1 + \alpha \ln h_i^* + \eta \ln E[h_i^*] + \varepsilon_j, \tag{7}$$

where $c_1 = \ln \beta - \eta \ln \left(\dfrac{\eta c_0}{1 + \eta} \right)$. The theory predicts that there are positive spillovers from average education since $\eta > 0$.

Up to this point, the model cannot explain why spillovers might be of different sizes for different subsamples. Suppose that the economy consisted

[4] We assume the uniform distribution in order to present a closed-form solution. As long as the expectation of g_i exists, however, the result that individual wages depend on average human capital still holds.

[5] At the time firms purchase capital, they are uncertain about the human capital level of the worker they will randomly meet and about their random productivity shock.

[6] As a technical matter, it is necessary to restrict α and Γ such that expected profits are non-negative; otherwise, firms would not remain in business. If $\alpha > 0.5$ and Γ is small, the non-negative expected-profits condition is satisfied.

of M subsamples, each of which had a different value of Γ, Γ_m.[7] Prior to the random meeting, the firm does not know from which subsample its worker will be drawn. Let k_m^* be the Nash equilibrium capital choice for a firm that knew it was sampling from subsample m. The expression for k_m^* is the same as that for k^*, but substituting Γ_m for Γ in the definition of η and using the resulting value (called η_m, say) in the expression for k^*. Let p_m be the population proportion of subsample m in the whole economy. It follows that $k^* = \Sigma_m p_m k_m^*$ and human capital is chosen by workers based on k^*, since they make this choice before they match with an employer. It is possible to show that there exists a function $V(\eta_1, \eta_2, ...\eta_M)$ that satisfies the following relationship:

$$\ln k^* = \frac{E[h_i^*]}{(1-\alpha)} \frac{1}{V(\eta_1, \eta_2, ...\eta_M)}. \tag{8}$$

Substituting (8) into (6), the new wage equation contains a term for expected human capital that depends on η_i. Further assumptions on the parameters[8] are needed to show that $\frac{\partial V}{\partial \eta_m} < 0$, making spillover effects positive. The wage equation we estimate can be viewed as a linear approximation to the one derived in the extension of the model to subsamples.

3. The Data

There are two data sources for this paper. Individual-level data come from the 1979–2000 waves of the geocoded National Longitudinal Survey of Youth, 1979 (NLSY79). State-level data are from the March Current Population Survey (CPS). This section describes these two sources in greater detail. Table 1 provides the sample means for the variables we use.

3.1 The NLSY79

Using information from household screening interviews in 1978, the NLSY79 includes a nationally representative cross-section plus oversamples of blacks, Hispanics, and poor whites, and a subsample of people serving in the military.

[7] One can think of the different values of Γ_m as being reduced-form representations of the different constraints faced by different groups in society. Since the Γ_m are exogenous in our model, we do not consider the origin of these differences.

[8] A sufficient condition is $\ln k > \ln(\alpha\beta)$.

We do not use the military subsample because we do not believe that our theoretical model is an adequate description of the wage formation process for people in the military. For similar reasons, we exclude a small sample of residents who describe themselves as self-employed. The geocode augmentation is a confidential file that includes detailed geographical information on respondents. These data connect the respondent to their labour market's human-capital proxy.

The NLSY79 personal variables used in the empirical model are hourly wages; population group; age; gender; tenure; married, spouse present; urban status; and health limitations. The NLSY's definition of the three population groups (whites, blacks, and Hispanics) is strictly non-overlapping.[9] We also use the following human capital variables from the NLSY79: educational attainment, mother's highest grade completed, father's highest grade completed, expected highest grade completed, reason for leaving school, and Armed Forces Qualification Test (AFQT) score.[10]

We exclude observations for respondents who usually worked less than 35 hours per week. For all remaining respondents, we observe hourly wage data that the NLSY79 compiles from data on earnings and hours worked. Specifically, one's hourly wage is the sum of all earnings from work (e.g., wages, tips, and bonuses) divided by usual hours worked per week. We create real wages using the national consumer price index-urban (CPI-U). Thus, we interpret changes in real wages as changes in purchasing power. To the best of our knowledge, we are the first in the spillover literature to use real wages in place of nominal ones. Using the real wage is necessary in panel regressions, since the nominal wage inherits the non-stationarity of the price level.[11] We omit observations in which nominal wages are below the then

[9] Population group data were collected during the household screening interviews in 1978. Hispanics were either self-identified as Hispanic or (i) they identified themselves as Filipino or Portuguese, (ii) they reported speaking Spanish at home as a child, and (iii) they had a Spanish surname identified by the census list of Spanish surnames. Blacks were identified as black by the interviewer and are non-Hispanic. Whites were defined as non-black, non-Hispanic (NLSY 2002).

[10] Most of the NLSY79 respondents completed the AFQT, but this does not indicate an intention to enter the armed forces. The AFQT score is derived from specific sections of the Armed Services Vocational Aptitude Battery (ASVAB).

[11] In panel regressions with few observations in the time dimension, standard tests for stationarity are impossible. We feel that the real wage is more likely to be stationary than the nominal one. In case the real wage is non-stationary, we compute regressions in differences and find results almost identical to those we obtain in levels.

prevailing national minimum wage. We also delete observations with real wages above one hundred dollars. We investigate all individuals whose real hourly wage exceeded fifty dollars in real earnings per hour in any sample year. Our deletion decisions are based on the earnings profile, education, and occupation of these respondents.[12] We analyze the population group, gender, and college completion status of the excluded wage outliers and the sample we keep; they are substantially similar. Thus, we feel confident that these sample deletions do not bias our results.

We use the following proxies for individual human capital: education, labour market experience, tenure, and AFQT score. Educational attainment data come from answers to the question: "What is the highest grade you completed?" We compute potential labour market experience as age minus highest grade completed minus six. The AFQT score "is a general measure of trainability" (NLSY 2002) and proxies for innate ability; the score is age-adjusted. "Married, spouse present" is a dichotomous variable often included in Mincerian regressions to proxy for a number of unobservable factors at the individual level that are related to productivity. An alternative justification for its inclusion is that respondents with a spouse can split home production responsibilities, which leaves more time for work. The NLSY79 also has a dichotomous variable for health limitations, which identifies respondents with health problems that restrict their ability to work.

All NLSY79 respondents were born between 1957 and 1964. We drop individuals in a given year who reported attending either an elementary, middle, or high school or college since the last interview. These individuals are not part of the labour force. We drop any observation where data are missing for one or more of our variables. Because we consider the decision to work an endogenous choice, we fit a sample selection equation. For the purposes of this equation, our sample has 9,306 individuals and 104,331 person-years. For the wage equation, we drop observations for which the respondent reported an activity during the survey week other than "working" (e.g., unemployed, keeping house, unable to work); observations with a non-positive wage are also dropped. After these deletions, the sample for the wage equation consists of 7,344 individuals and 65,882 person-years.

[12] The Center for Human Resource Research, which compiles raw NLSY79 data into hourly wage, does not correct abnormally high or low wage outcomes. These wages could be a result of an interviewer's inputting mistake or respondent mistruth, or they could be correct. For this reason, we felt it necessary to investigate these cases individually.

3.2 The CPS

We consider the respondent's state as the surrounding labour market. We proxy for state-level human capital using the percentage of college graduates for each state and Washington, D.C.[13] Data for the percentage of college graduates come from the CPS. In calculating the percentage of college graduates, we drop CPS respondents younger than 25 to eliminate those currently in school. We also drop those older than 65, since they are likely to be out of the labour force. Prior to 1992, the CPS collected data on years of education instead of data on completed degrees. For those years, we consider respondents with at least four years of college to be college graduates.

Rauch (1993), Rudd (2000), Acemoglu and Angrist (2000), and Ciccone and Peri (2002) use the mean highest grade completed as the proxy for human capital. We feel that the percentage of college graduates is preferable for two reasons: multicollinearity and the distribution of educational attainment. Multicollinearity among the regressors increases the difficulty of separating their effect on wages. By construction, individual highest grade completed and its statewide mean are highly correlated.[14] There is a smaller correlation, however, between the percentage of college graduates and individual highest grade completed. An increase in an individual's highest grade completed only increases the statewide percentage of college graduates in the case where the increase is from 15 (completion of junior year in college) to 16 (a college graduate). Most other gains in schooling increase the likelihood of a higher percentage of college graduates in the future, but do not affect this variable contemporaneously. Conditioning-number tests for multicollinearity, recommended by Belsley, Kuh, and Welsch (1980), confirm this logic. These tests suggest that multicollinearity is present between our other regressors and both human capital proxies; the mean highest grade completed exhibits a higher degree of multicollinearity than the statewide percentage of college graduates.

In addition, there is evidence that the distribution of highest grade completed is both bimodal and skewed to the right. In general, the mean is a poor summary statistic for skewed distributions, especially bimodal ones. The bi-

[13] Moretti (2004a) also uses percentage of college graduates as his proxy for local human capital, although he uses MSAs as his local labour market (instead of states).

[14] This is true even when the individual data are taken from the NLSY79 and the statewide averages are based on the CPS, to the extent that both data sets are random samples from the same data-generating process.

modal aspect of the distribution is likely a by-product of discrete jumps at high school and college graduations, known as "sheepskin effects." Studies of the wage-schooling profile suggest returns to education are not linear in years of schooling attainment. Regardless of their cause—discrete jumps in skill, signalling value of graduation, etc.—sheepskin effects influence education decisions, causing people to cluster at a terminal high school degree and a terminal college degree.[15] We use percentage of college graduates as our proxy for statewide human capital based on these considerations.

Ideally, one would want to introduce some measure of school quality, since college degrees are heterogeneous. Unfortunately, we cannot disaggregate data on statewide college graduation by the quality of that education.

4. The Empirical Model

To test equation (7), we need to estimate a Mincerian regression. Section 2 showed that both h and $E[h]$ are endogenous.[16] We estimate first-stage regressions and use instrumental variables for individual and average human capital to obtain consistent estimates of the marginal effect of h and $E[h]$ on wages.[17]

4.1 First-stage regressions

Because the NLSY79 is a panel, individual highest grade completed is not constant over time for all individuals. For example, some respondents could have two "spells" in the labour force if they did not go directly to college from high school, causing variation across time in their educational attainment. We must estimate separate first-stage models for those with interrupted-schooling behaviour and those without. In addition, there are two types of uninterrupted-schooling individuals: those who completed their education prior to the start of the panel in 1979, and those who completed it afterwards. Therefore, we use three separate regressions to instrument for the

[15] We address the endogeneity of educational attainment by using instrumental variables.

[16] Ordinary least squares regressions and Hausman tests confirm that h and $E[h]$ are also endogenous in our data.

[17] Acemoglu and Angrist (2000) show that it is necessary to treat both statewide and individual human capital as endogenous to guarantee consistent estimates. Moretti (2004a) uses instrumental variables for average human capital. Our instrumental variables section borrows from these papers to some extent.

highest grade completed. For the uninterrupted-schooling individuals, whose education was completed before 1979, we fit the following regression:

$$\begin{aligned} hgc_i &= \psi_0 + \psi_1 mhgc_i + \psi_2 \mathit{fhgc}_i + \psi_3 \mathit{afqt}_i + \psi_4 \mathit{volpos}_i + \psi_5 \mathit{volneg}_i + \\ &\quad \sum_{r=6}^{8} \psi_r pgg_{ir} + \varepsilon_{1i}, \end{aligned} \tag{9}$$

where *hgc* is highest grade completed; *mhgc* and *fhgc* are, respectively, mother's and father's highest grade completed; *afqt* is the individual's score on the Armed Forces Qualification Test; and pgg_i are three dummy variables for population group and gender—specifically, blacks, Hispanics, and females (white males are the omitted category). The two dummy variables, *volpos* and *volneg*, indicate that the respondent left school voluntarily for "positive" reasons and "negative" reasons, respectively.[18] This regression allows for the possibility that human capital acquisition is affected by family background, intelligence, and tendencies to acquire human capital that may differ by gender and population group.[19] These non-time-varying factors do not enter the second-stage wage regression, since they are absorbed by an individual-level fixed effect. We use the reason why an individual left school before 1979—positive, negative, or involuntary—as the instrumental variable for highest grade completed. The NLSY asked why a respondent left school only if the respondent was not in school in 1979. These "reason-for-leaving" variables affect education decisions but are not likely to be correlated with the unexplained portion of their wage at a later date. We estimate the following regression for uninterrupted-schooling individuals who left school in 1979 or later:

[18] The omitted category is "left school involuntarily." These three dummy variables are constructed from 12 separate responses to the question "What is the main reason you left [school] at that time?" Details regarding which responses are positive, negative, and involuntary are provided in the appendix.

[19] There is evidence that these effects of differential tendencies to acquire human capital by population group are diminishing over time. In 1940 (based on census data), on average, whites had more than three years more education than blacks or Hispanics. By 2000, that gap had closed to less than two years for Hispanics and less than one year for blacks. We make no attempt to explain the source of these tendencies.

$$hgc_i = \chi_0 + \chi_1 mhgc_i + \chi_2 fhgc_i + \chi_3 afqt_i + \chi_4 E_{1979}\left[hgc_i\right] + \sum_{r=5}^{7} \psi_r pgg_{ir} + \varepsilon_{2i}, \tag{10}$$

where $E_{1979}[hgc_i]$ designates the highest grade that individuals expect to complete. Expected education is the instrument for the uninterrupted-schoolers who finished school after 1979. The NLSY asked this question only in 1979. We feel that individuals' expectations of their highest grade completed (formed in 1979) is unlikely to be correlated with the unexplained portion of their wages several years later, when they have left school and are earning a wage.

For interrupted-schooling individuals, whose highest grade completed changes over time, we fit the following regression:

$$\begin{aligned} hgc_{it} \;=\; & \phi_0 + \phi_1 mhgc_i + \phi_2 fhgc_i + \phi_3 afqt_i + \phi_4 E_{1979}\left[hgc_i\right] + \\ & \sum_{r=5}^{7} \phi_r pgg_{ir} + \phi_8 age_{it} + \varepsilon_{3it}, \end{aligned} \tag{11}$$

where *age* is the age of the individual at time t. In our wage regression, we combine all of these types of individuals, using the fitted values from (9), (10), and (11). This combination of fitted values has the potential to create heteroscedastic errors. In all our wage regressions, we correct the errors for this and other sources of heteroscedasticity; our discussion is below.

We also need instrumental variables for statewide college share, which is our proxy for statewide human capital. It is plausible that college share and wages are affected by variables unobserved in the wage equation (7). For example, high wages in a given state may increase demand for schooling in that state, which increases the level of human capital. Alternatively, high-skilled workers may migrate to states with high wages. This endogeneity biases the coefficient on the percentage of college graduates in the wage equation, which is our estimate of the net effect of statewide human capital on wages. Acemoglu and Angrist (2000), Ciccone and Peri (2002), Rudd (2000), and Moretti (2004a) account for this potential endogeneity in their models, although they use different specifications. We follow Moretti (2004a) and use an age distribution from the 1970 census in the instrument for state human capital. We construct

$$iv_{st} = \sum_p f_{ps} nmcol_{pt}, \tag{12}$$

where p indexes age strata in years (from 25 to 65, inclusive), f_{ps} denotes the fraction of the population of state s that was in stratum p in 1970, and $nmcol_{pt}$ indicates the nationwide fraction of college graduates in stratum p at time t.[20] Use of the 1970 age distribution ensures that this variable is exogenous to the wage equation error term, since there is no reason to believe that the 1970 age distribution affects the unexplained component of wages between 1979 and 2000. The nationwide fraction of college graduates is also orthogonal to the wage equation error, for two reasons. First, nationwide college share is not affected by non-random migration within the United States. Second, since there are 51 "states," none of which is large relative to the whole, endogenous changes in college share in one state barely change the national average. As with Moretti (2004a), we argue that iv_{st} is correlated with the endogenous statewide human capital variable via the national upward trend in education over the sample period. States with "young" populations in 1970 are likely to have the largest share of college graduates in the years after 1979 (i.e., during the sample). We account for the endogeneity of $mcol$ in the following fixed-effects regression:

$$mcol_{st} = \xi_1 iv_{st} + \xi_{2s} + \varepsilon_{4st}. \tag{13}$$

To simplify the interpretation of the results, we omit the constant term and include all of the state-level fixed effects, instead of omitting one of the fixed effects and including a constant term.

Finally, although it is not our purpose to model labour supply, we need to control for the potential bias introduced by respondents in the sample who are not working. This is particularly problematic, since we estimate spillover effects across population group, gender, and college completion status, creating subsamples that differ substantially in their labour supply. We fit a standard employment probit model to the entire sample, and use the fitted values from this as the Heckman regressor in the wage regression. Our model is:

[20] Moretti's (2004a) instrumental variable regression uses the first difference of $nmcol_{pt}$.

$$\Pr\left(w_{it} > 0\right) = \phi\left(\begin{array}{c} \mu_0 + \sum_{r=1}^{2} \mu_r edu_{rit} + \mu_3 unemp_{s,t-1} + \\ \sum_{r=4}^{6} \mu_r pgg_{rit} + \sum_{r=7}^{9} \mu_r control_{rit} + \varepsilon_{5it} \end{array}\right), \tag{14}$$

where edu_1 denotes the possession of a high school diploma, edu_2 denotes the possession of a college degree, $unemp$ is the lagged state unemployment rate, pgg are the population group and gender variables used in equations (8)–(10) and the controls are, respectively, urban resident, health limitations, and married, spouse present. Since this is a probit regression, ϕ is the standard Gaussian density function. Our identifying instrument is the lagged state unemployment rate, which we expect to be orthogonal to the unexplained portion of an individual's wage. With the fitted values for highest grade completed and average statewide college completion, as well as the Heckman term from the probit regression, we are ready to consider the specification of the wage regression.

4.2 Second-stage regressions

We begin by fitting a standard Mincerian regression, inspired by (7):

$$\begin{aligned} \ln w_{it} &= \pi_0 + \pi_{1i} + \pi_2 \widehat{hgc}_{it} + \pi_3 \widehat{hgc}_{it} \cdot cg_{it} \\ &\quad + \pi_4 \sum_{s=1}^{51} I\left(i \in s, t\right) \cdot \widehat{mcol}_{st} + X_{it}' \Pi + \varepsilon_{6it}, \end{aligned} \tag{15}$$

where hats above variables denote the fitted values from the first-stage regressions and I is an indicator variable that equals one if person i lives in state s at time t, and zero otherwise. Mincerian regressions typically include a dummy variable for college graduation; the estimated coefficient captures the sheepskin effect. But for respondents who did not interrupt their schooling, their college completion dummy is perfectly correlated with their fixed effect. To avoid this problem but still account for sheepskin effects, we compute two individual human capital variables using the estimates from the first-stage individual human capital regression: $\widehat{hgc}$ is the fitted value from the first stage and cg is a dummy variable equalling one if $\widehat{hgc} \geq 16$. As shown above, for those with a college degree and higher, the marginal benefit of an additional year of schooling is the sum of the coefficients of $\widehat{hgc}$ and

$\widehat{hgc} \cdot cg$.[21] There are seven other control variables, standard to Mincerian regressions: work experience and its square, tenure at the individual's current job and its square, whether the individual has health limitations, whether the individual is married with spouse present, and whether the individual lives in an urban area. Although less common in the Mincerian literature, we also include two "macroeconomic" control variables, log gross state product (GSP) and a time trend. GSP accounts for the variation in growth across states. Assuming that statewide economic growth and human capital are positively correlated, omitting GSP from the model will bias the estimate of the spillover effect, since economic growth likely affects statewide education (via migration) and individual wages.[22] A time trend is not usually included in panel regressions, but log GSP is trend stationary.[23] Including a trend-stationary regressor without a time trend makes the distribution of the regression error non-stationary. If the regression error is non-stationary, its distribution may not converge to the normal distribution—a fact that complicates inference. These nine control variables are included in the X_{it}' row vector in equation (15).

We estimate another version of (15) to account for spillover differences across subsamples:

$$\begin{aligned}\ln w_{it} &= \theta_0 + \theta_{1i} + \theta_2 \widehat{hgc}_{it} + \theta_3 \widehat{hgc}_{it} \cdot cg_{it} \\ &\quad + \sum_{m=4}^{M+3} I\left(i \in m\right) \cdot \theta_m \sum_{s=1}^{51} I\left(i \in s, t\right) \cdot \widehat{mcol}_{st} + X_{it}'\Theta + \varepsilon_{7it}, \quad (16)\end{aligned}$$

[21] Because of multicollinearity problems, it is not possible to include $\widehat{hgc}$, cg, and their product in the same regressions. We encounter similar problems if we include interaction terms of $\widehat{mcol}$ with dummies for our 12 subsamples as well as those dummies themselves. As a robustness check (in both cases), we try including only the linear terms and dropping the interaction terms; for variables included in both regressions, the results are broadly similar.

[22] Nominal GSP is the sum of the final output of a state multiplied by the prices of those outputs. We create real GSP by dividing by the national CPI-U. To the extent that there are interstate variations in prices, they remain in our variable, log real GSP. Since wage bargaining may be affected by the local price level, log real GSP may be correlated with the error term of the wage equation, thus biasing the results. The simple correlation coefficient between log real GSP and the residuals is about 2.5 per cent, which suggests that this potential source of endogeneity is not a problem in our sample.

[23] Stationarity test results confirming that log GSP is trend-stationary in this sample are available upon request.

where m indexes subsamples. We create 12 subsamples by dividing the sample along three axes: college educated versus not college educated, population groups (whites, blacks, and Hispanics), and gender. We perform t-tests to determine that the θ_m are different from one another.

For regressions (15) and (16), we use standard errors corrected for general forms of heteroscedasticity, as in White (1980).[24] Moulton (1986) provides an alternative methodology for addressing heteroscedasticity, known as clustered errors. He considers the case where the regression errors have common variance but are correlated within groups and uncorrelated across groups. Acemoglu and Angrist (2000) use state-time clustering. For our model, the logical cluster is at the state-time level, since individual-level shocks are accounted for in the fixed effect. If we clustered on the state of residence alone, it would be similar to estimating a state fixed effect.[25] Clustering on time alone assumes that all states experience a random shock in a given year, which is inconsistent with our view that the state is the relevant local labour market in which spillovers may occur.[26]

We reject the use of clustered errors for four principal reasons. First, the fixed effect absorbs much of the state-level shocks. This is particularly true for non-migrants, whose state of residence does not change throughout the sample. Second, we are unaware of any paper using clustered errors in a panel data context, and thus their statistical properties are unknown. Third, we test for residual correlation by drawing random samples of pairs of residuals from the second-stage regression. Each member of each pair came from the same state-time combination. Simple correlation coefficients averaged over 100 samples of 100 pairs each are about 2 per cent. This diagnostic technique suggests that clustered errors are not necessary for this data set. Fourth, the results using clustering are broadly similar to those using White (1980) errors.

[24]We also check the second-stage residuals of the three types of individuals in our sample (those who interrupted their schooling, those who did not and finished after 1979, and those who had completed their schooling before 1979), to determine whether their distributions are the same; we find that they are.

[25]We tried a version of our model with state fixed effects, but found that most of them were insignificant.

[26]Estimating a fixed-effects regression with clustered errors is an option using STATA's "areg" command.

5. Results

5.1 First-stage regression results

Table 2 provides results of the first-stage individual human capital regressions for those who did not interrupt their schooling and left school prior to 1979. As expected, parents' educational outcomes positively affect the child's highest grade completed. There are positive coefficients for blacks and Hispanics (whites are the omitted category), likely caused by faster growth in education for blacks and Hispanics than for whites, which produces a larger difference between each generation's education. The instrumental variables for this group are dummy variables for the reason the respondent voluntarily left school, condensed into positive reasons and negative reasons. Leaving school involuntarily is the omitted category. Those who leave school for positive reasons, such as graduation, stay in school longer. Similarly, those who leave school for negative reasons obtain less education. Both of these instruments are statistically significant.

Table 3 reports results of the first-stage individual human capital regressions for those who did not interrupt their schooling patterns but left school after 1979. Parents' educational outcomes and AFQT have positive and statistically significant effects on education. The coefficient for blacks is positive and statistically significant, which is again a likely by-product of faster growth in education than for whites. The instrumental variable for this group is the expectation of highest grade, formed in 1979, which has a positive and significant effect on the actual educational outcome.

Table 4 provides results of the first-stage individual human capital regressions for those with interrupted schooling patterns. Since this subset of the sample has time-varying educational outcomes, we include a time-varying regressor: age. Among this group, highest grade completed is positively related to parents' educational outcomes, AFQT, and whether the respondent is in the black population group. Highest grade completed is negatively related to whether the respondent is in the Hispanic population group, indicating that Hispanics that interrupt their schooling have slower education growth than whites who do likewise. The instrumental variable for this group is again the expectation of highest grade completed, formed in 1979, which has a positive and significant effect on the actual educational outcome.

Table 5 reports results of the employment selection equation. All variables have the expected sign: positive effects from education, marriage, and urban status; negative effects from blacks, females, and Hispanics; and health limitations. The instrumental variable is the lagged unemployment rate, which has a negative effect on employment.

For brevity, we suppress the results of the first-stage regression of statewide college share on Moretti's (2004a) exogenous age structure variable and state controls (i.e., equation (13)). The coefficient estimate for the age structure variable is 0.96 with a t-statistic of 43.71, which suggests that the instrument is highly correlated with statewide college share.

5.2 Second-stage regression results

Table 6 provides results of the second stage using predicted values for hgc and $mcol$. We find evidence of spillover effects. A 1 percentage point increase in college share increases individual wages by about 3 per cent. Moretti (2004a) and Acemoglu and Angrist (2000) report smaller positive spillover estimates, although only Moretti's are statistically significant. Prior to college graduation, the marginal return to schooling is 7 per cent, which is consistent with other Mincerian estimates. We find the familiar concave marginal effects of experience and tenure, and expected signs for health limitations; married, spouse present; urban status; and GSP. The time-trend variable is not statistically different from zero, which suggests that the time trends in other regressors and in log wages "cancel out" the time trend in log GSP.

5.3 Second-stage results: the subsamples model

The model in Table 6 assumes that all subsamples receive the same spillover effect. In the extant spillover literature, only Moretti (2004a, b) considers the possibility of group-specific spillovers, finding that college graduates receive a smaller spillover effect than non-college graduates. There could also be differences among other observable dimensions, such as population group and gender, perhaps caused by discrimination or differences in the occupation distribution of each subsample. Our model does not attempt to address these issues. Table 7 provides results for the model with separate spillover effects for each of our 12 subsamples. These results show that there is variation in the spillover effect across our subsamples. The spillover effect ranges from 1.5 per cent (non-college-educated black males) to over 5 per cent (college-educated white males) across our 12 subsamples. The other variables in the model have coefficients that are similar to those in Table 4. The time-trend variable is again negative, but now statistically significant.

5.4 Comparison of spillover effects

We perform t-tests to find statistical differences between pairs of spillover estimates.[27] If $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimated spillover coefficients with associated variances σ_1^2 and σ_2^2, the statistic $\dfrac{|\beta_1 - \beta_2|}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ has an asymptotic standard normal distribution, provided that the subgroups for which $\hat{\beta}_1$ and $\hat{\beta}_2$ are estimated are non-overlapping. We exploit the non-overlapping feature of our subgroups to make proper inferences.[28]

The first type of difference we identify is the "college effect," or differences in the spillover coefficients between those who have a college degree and those who do not. Economic theory suggests that this effect may be positive or negative. Supply effects imply that an increase in college graduates will reduce the spillover return to obtaining a college degree; this is the effect found by Moretti (2004a). Alternatively, knowledge sharing between workers may cause increasing returns to statewide human capital accumulation. We find a positive college effect for two of our six comparison groups: white males and black males. In coarser cuts of the data, there is no statistical difference between the spillover for college graduates and non-college graduates. We believe that treating both *hgc* and *mcol* as endogenous in addition to our macroeconomic control variables accounts for the difference between Moretti's (2004a) estimates and ours.

We find gender effects for three of the six comparisons: whites who are college educated, Hispanics who are college educated, and whites who do not have a college degree. In each comparison, males receive a larger spillover benefit than females. Interestingly, we find no evidence of male-female spillover differences when the data are cut by gender only, and we ignore spillover differences across population group and college completion. These results suggest that cutting the data by gender only does not tell the full story.

[27] The t-test for differences does not rely on the individual estimates being significantly different from zero. Significant differences can be ascertained as long as estimates of the coefficients and their variances are consistent.

[28] In the NLSY79 data, whites, blacks, and Hispanics are non-overlapping categories, which is not the case in Census Bureau or CPS data.

Between whites and blacks, we find effects in two of the four comparisons: men who are college educated, and men who do not have a college degree. In each case, whites receive a larger spillover effect than blacks. Between blacks and Hispanics, we find population group effects for two of the four comparisons: men who are college educated and men who are not college educated. Hispanics receive a larger spillover effect than blacks in both. Between whites and Hispanics, only white men who are college educated receive a statistically different spillover than Hispanic women who are college educated. All of the significantly different population group and college effects are among males. Based on our explanation for spillovers, this empirical regularity suggests that women may be more likely to share information across population groups and educational status.

We identify three types of spillover differentials (population group, gender, and educational attainment). Can we explain these differentials using observables? One explanation for differing spillover effects is in the types of jobs held by different subsamples. Interaction among workers facilitates information sharing, which affects the spillover coefficient. Presumably, the amount of worker interaction differs across jobs because of differences in the production process. Our interpretation of spillovers also requires some degree of flexibility in the production process to incorporate new information. For example, the regimented production process of assembly-line workers makes it unlikely that information sharing will affect productivity. We proxy for these job-specific differences using 12 broad occupation groups from the 1970 Census Bureau classification. We then run separate second-stage regressions for each occupation group to test for spillover differences. Table 8 reports the results. Statistically significant point estimates range from -0.106 to 0.062. However, t-tests show that the differences in the spillover coefficient are not statistically different from one another in nearly all of the pairwise occupation comparisons.

To determine the explanatory power of occupational differences, we use the spillover coefficients for each occupation and each subsample's occupation distribution. Occupation differences can only predict the spillover differences across subsamples if there are systematic differences in occupation across subsamples. We examine this empirical question and begin by computing the empirical distribution of participation in 12 broad occupation groups by subsample. Let π_m be the empirical distribution of occupations for subsample m. We assume that this distribution is a consistent estimator of the true distribution. We place the spillover coefficients for each occupation reported in Table 8 in μ_o. Let μ_m be the vector of coefficients on $\widehat{mcol}$ interacted with our 12 subsamples stratified by population group, gender, and college completion status. We test whether the product $\mu_o' \pi_m$ is a good predictor of μ_m by linear regression, finding an adjusted R^2 of about 0.2. Thus, occupation has limited predictive power for spillover effects.

It is also possible that imperfect knowledge sharing drives the differences in spillover effects. Prejudices against, and favouritism towards, a specific subsample disrupt the flow of information among workers, causing productivity differences. While it would be ideal to test this hypothesis directly, it is difficult to do so for two reasons. First, our information-transmission explanation refutes Rauch's contention that statewide human capital is a public good, since imperfect knowledge sharing implies some degree of excludability. It is more difficult to measure excludability (i.e., the absence of communication or spillovers) than it is to measure inclusiveness or the presence of spillovers. Second, although there is a substantial literature on labour market discrimination beginning with Becker (1957), this literature is not easily applied to spillover discrimination. For these reasons, we do not test for discrimination. Rather, we say that 20 per cent of the spillover differences is attributable to occupational choice, which leaves 80 per cent of the difference unexplained.

6. Conclusions

There is no current consensus regarding the impact of an increase in average human capital on individual wages. Economic theory suggests that human capital accumulation in a labour market affects individual wages in both directions. An increase in the supply of highly skilled workers decreases the wages of all highly skilled workers. But human capital accumulation also increases the stock of knowledge. Workers share some of this information, which raises their productivity and wages. Identification of the net effect of statewide human capital accumulation on individual wages presents several challenges. Proxies for statewide human capital are almost certain to be endogenous in the wage equation, because of factors such as non-random migration or local-specific growth shocks. Acemoglu and Angrist (2000) show that it is also necessary to treat individual human capital as endogenous if the average human capital variable is constructed from data on individual human capital. A reliable estimate of spillover effects is an important factor in decisions on public education funding, since spillovers represent the externality that all citizens enjoy of marginal increases in statewide human capital.

Our contributions to the literature are two-fold. First, we find positive effects from statewide human capital accumulation; i.e., the knowledge spillover effect dominates the supply effect. A 1 percentage point increase in college graduates in a given state increases individual wages by about 3 per cent. We find that all workers receive spillovers, regardless of whether we consider an aggregate sample or various disaggregated subsamples. Second, spillovers are not equal for all groups of workers. In general, spillovers are largest for whites and smallest for blacks, larger for the college educated than for those who do not have a college degree, and larger for men than they are for women. We estimate that roughly 20 per cent of these spillover differences are attributable to differences in the distribution of broad occupation groups across subsamples. Our theoretical and empirical models are neither able to support nor to negate an explanation of spillover differences based on prejudice or favouritism. Our results suggest future spillover research should account for these differences in spillovers across groups. In addition, determining the cause of spillover differences across population groups, gender, or education level may provide insight into how the spillover externality works and into policies to encourage spillovers.

References

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

Acemoglu, D. and J. Angrist. 2000. "How Large Are the Social Returns to Education: Evidence from Compulsory Schooling Laws." *National Bureau of Economic Research Macroeconomics Annual* 15(1): 9–60.

Becker, G. 1957. *The Economics of Discrimination.* Chicago, IL: University of Chicago Press.

Belsley, D., E. Kuh and R. Welsch. 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* New York, NY: John Wiley and Sons.

Ciccone, A. and G. Peri. 2002. "Identifying Human Capital Externalities: Theory with an Application to U.S. Cities." University of California. Photocopy.

Friedman, M. 1962. *Capitalism and Freedom.* Chicago, IL: University of Chicago Press.

Glaeser, E. 1999. "Learning in Cities." *Journal of Urban Economics* 46(2): 254–77.

Lucas, R. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22(1): 3–42.

Marshall, A. 1890. *Principles of Economics.* New York, NY: Macmillan.

Mincer, J. 1974. *Schooling, Experience and Earnings.* New York, NY: Columbia University Press.

Moretti, E. 2004a. "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics* 121(1–2): 175–212.

Moretti, E. 2004b. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review* 94(3): 656–90.

Moulton, B. 1986. "Random Group Effects and the Precision of Regression Estimates." *Journal of Econometrics* 32(3): 385–97.

NLSY. 2002. *A guide to the 1979–2000 National Longitudinal Survey of Youth Data*. Columbus, OH: Center for Human Resource Research, The Ohio State University.

Rauch, J. 1993. "Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities." *Journal of Urban Economics* 34(3): 380–400.

Rudd, J. 2000. "Empirical Evidence on Human Capital Spillovers." Board of Governors of the Federal Reserve System, Finance and Economics Discussion Paper No. 2000-46.

White, H. 1980. "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity." *Econometrica* 48(4): 817–38.

Table 1: The Data

Variable	Mean	Standard deviation
Log wage	1.91	0.48
Highest grade completed: hat	12.90	1.85
Mean college: hat	22.57	2.21
Experience	9.51	5.40
Tenure	3.61	3.89
Health limitations	0.04	0.20
Married, spouse present	0.52	0.50
Urban	0.78	0.42
Log gross state product	11.78	0.88
Expected highest grade completed	13.83	2.48
Mother's highest grade completed	11.06	3.04
Father's highest grade completed	10.97	3.83
Age	28.48	5.43
Female	0.48	0.50
Black	0.23	0.42
Hispanic	0.15	0.36
Armed Forces Qualification Test	52.27	27.86

Table 2: First Stage: Non-Interrupted, Left before 1979

Variable	Estimate	Variable	Estimate
Mother's highest grade completed	0.10 (15.77)	Black	0.80 (20.00)
Father's highest grade completed	0.05 (5.25)	Female	0.28 (8.10)
Armed Forces Qualification Test	0.02 (28.28)	Hispanic	0.08 (2.48)
Left voluntarily, positive reasons	0.74 (16.78)	Left voluntarily, negative reasons	−0.59 (11.00)
		Constant	8.20 (153.25)
N	2204	R^2	0.4517

Table 3: First Stage: Non-Interrupted, Left after 1979

Variable	Estimate	Variable	Estimate
Mother's highest grade completed	0.05 (4.22)	Black	0.52 (10.69)
Father's highest grade completed	0.04 (4.88)	Female	0.07 (1.08)
Armed Forces Qualification Test	0.03 (48.38)	Hispanic	0.16 (0.80)
Expected highest grade completed	0.53 (25.09)	Constant	3.08 (15.09)
N	3273	R^2	0.5897

Table 4: First Stage: Interrupted

Variable	Estimate	Variable	Estimate
Mother's highest grade completed	0.02 (3.18)	Black	0.29 (8.69)
Father's highest grade completed	0.04 (9.08)	Female	0.28 (11.37)
Armed Forces Qualification Test	0.03 (42.64)	Hispanic	−0.07 (−1.76)
Age	0.12 (52.12)	Expected grade completed	0.39 (59.95)
		Constant	2.13 (20.11)
N	18015	R^2	0.4940

Table 5: Probit Results

Variable	Estimate	Variable	Estimate
High school graduate	0.67 (58.04)	Black	−0.21 (−17.95)
College graduate	0.27 (16.26)	Female	−0.65 (−62.91)
Lagged state unemployment	−0.02 (−10.75)	Hispanic	−0.03 (−2.30)
Health limitations	−0.61 (−34.93)	Married, spouse present	0.02 (1.65)
Urban	0.04 (3.33)	Constant	1.01 (52.22)
N	104331	Pseudo-R^2	0.1076

Table 6: Second-Stage Results: Single Population

Variable	Estimate	Variable	Estimate
Highest grade completed: hat	0.070 (14.68)	Health limitations	−0.025 (−4.03)
Highest grade completed: hat ·College graduate	0.002 (2.65)	Married, spouse present	0.024 (7.76)
Mean college: hat	0.032 (9.41)	Urban	0.016 (3.44)
Experience	0.020 (5.99)	Log gross state product	0.042 (12.18)
Experience^2	−0.0008 (−20.89)	Time	−0.003 (−1.07)
Tenure	0.036 (40.72)	Constant	−0.298 (−3.13)
Tenure^2	−0.0015 (−26.77)	Heckman term	−0.352 (−6.92)
N	65882	R^2	0.3116

Table 7: Second-Stage Results: 12 Subsamples

Variable	Estimate	Variable	Estimate
Highest grade completed: hat	0.068 (14.04)	Mean college: hat ·college, white, male	0.051 (14.35)
Highest grade completed: hat ·College graduate	0.001 (1.03)	Mean college: hat ·no college, white, male	0.037 (10.83)
Experience	0.034 (9.52)	Mean college: hat ·college, white, female	0.030 (8.63)
Experience^2	−0.0007 (−20.29)	Mean college: hat ·no college, white, female	0.021 (6.13)
Tenure	0.036 (41.01)	Mean college: hat ·college, Hispanic, male	0.039 (8.62)
Tenure^2	−0.0015 (−27.69)	Mean college: hat ·no college, Hispanic, male	0.029 (7.51)
Health limitations	−0.024 (−3.86)	Mean college: hat ·college, Hispanic, female	0.026 (6.21)
Married, spouse present	0.021 (6.75)	Mean college: hat ·no college, Hispanic, female	0.025 (6.36)
Urban	0.019 (4.13)	Mean college: hat ·college, black, male	0.027 (6.19)
Log gross state product	0.044 (12.80)	Mean college: hat ·no college, black, male	0.015 (3.95)
Time	−0.016 (−4.70)	Mean college: hat ·college, black, female	0.024 (6.05)
Constant	−0.204 (−2.14)	Mean college: hat ·no college, black, female	0.020 (5.28)
Heckman term	−0.333 (−6.46)		
N	65882	R^2	0.3107

Table 8: Regression Results by Occupation

Occupation	Spillover	Std. error	N
Professional/Technical workers	0.045	0.0090	10009
Managers/Administrators (non-farm)	0.062	0.0121	7227
Sales workers	0.054	0.0231	3120
Clerical and unskilled workers	0.034	0.0064	13848
Craftsmen	0.052	0.0103	7804
Operatives (non-transport)	0.028	0.0097	7243
Transport operatives	0.001	0.0198	2645
Labourers (non-farm)	0.033	0.0162	4165
Farmers/Farm managers	-0.375	0.2340	88
Farm labourers/Foremen	-0.106	0.0538	427
Service workers (non-private)	0.023	0.0099	8892
Private household workers	0.151	0.1723	187

Appendix

A.1 Data Abbreviations

afqt: score on the Armed Forces Qualification Test
edu_1: high school graduate
edu_2: college graduate
cg: dummy, one if the respondent is a college graduate
control: health limitations; married, spouse present; or urban
fhgc: father's highest grade completed
hgc: highest grade completed
mcol: the percentage of college graduates, by state of residence
mhgc: mother's highest grade completed
pgg: population group or gender dummy; i.e., blacks, females, or Hispanics
time: a time trend
unemp: state unemployment rate, annual
urban: respondent lives in an urban area

A.2 Variable Definitions: Dummies for Reasons for Leaving School

Voluntary positive: received degree, completed coursework, got married, home responsibilities, got good job, or entered military

Voluntary negative: pregnancy, poor grades/low ability, did not like school or moved away from school

Involuntary: expelled, suspended, school became too dangerous, financial difficulties

Respondents who did not list one of these 12 answers were deleted as missing data.

We consider the separation of these reasons for leaving school somewhat arbitrary and certainly subject to debate.

Bank of Canada Working Papers
Documents de travail de la Banque du Canada

Working papers are generally published in the language of the author, with an abstract in both official languages. ***Les documents de travail sont publiés généralement dans la langue utilisée par les auteurs; ils sont cependant précédés d'un résumé bilingue*.**

2005

2005-9	State Dependence in Fundamentals and Preferences Explains Risk-Aversion Puzzle	F. Chabi-Yo, R. Garcia, and E. Renault
2005-8	Recent Developments in Self-Employment in Canada	N. Kamhi and D. Leung
2005-7	Determinants of Borrowing Limits on Credit Cards	S. Dey and G. Mumy
2005-6	Monetary Policy under Model and Data-Parameter Uncertainty	G. Cateau
2005-5	Y a-t-il eu surinvestissement au Canada durant la seconde moitié des années 1990?	S. Martel
2005-4	State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation?	P.J. Klenow and O. Kryvtsov
2005-3	Pre-Bid Run-Ups Ahead of Canadian Takeovers: How Big Is the Problem?	M.R. King and M. Padalko
2005-2	The Stochastic Discount Factor: Extending the Volatility Bound and a New Approach to Portfolio Selection with Higher-Order Moments	F. Chabi-Yo, R. Garcia, and E. Renault
2005-1	Self-Enforcing Labour Contracts and the Dynamics Puzzle	C. Calmès

2004

2004-49	Trade Credit and Credit Rationing in Canadian Firms	R. Cunningham
2004-48	An Empirical Analysis of the Canadian Term Structure of Zero-Coupon Interest Rates	D.J. Bolder, G. Johnson, and A. Metzler
2004-47	The Monetary Origins of Asymmetric Information in International Equity Markets	G.H. Bauer and C. Vega
2004-46	Une approche éclectique d'estimation du PIB potentiel pour le Royaume-Uni	C. St-Arnaud
2004-45	Modelling the Evolution of Credit Spreads in the United States	S.M. Turnbull and J. Yang

Copies and a complete list of working papers are available from:
Pour obtenir des exemplaires et une liste complète des documents de travail, prière de s'adresser à :

Publications Distribution, Bank of Canada
234 Wellington Street, Ottawa, Ontario K1A 0G9
E-mail: publications@bankofcanada.ca
Web site: http://www.bankofcanada.ca

Diffusion des publications, Banque du Canada
234, rue Wellington, Ottawa (Ontario) K1A 0G9
Adresse électronique : publications@banqueducanada.ca
Site Web : http://www.banqueducanada.ca