April 1997

Interpreting the Coefficient of Schooling in the Human Capital Earnings Function

Barry R. Chiswick*

I. Introduction

The "human capital earnings function" (HCEF) has become a fundamental tool in research on earnings, wages and incomes in developed and developing economies.[1] It is an accepted procedure in litigation involving earnings, such as cases involving the value of lost earnings due to injury, death or discrimination (see, for example, Gastwirth, 1988 and Federal Judiciary Center, 1994). It is also frequently used to make educational policy decisions based on estimates of the rate of return from schooling (see, for example, Psacharopoulos and Mattson, 1996).

The basic feature of the HCEF is that it relates the natural logarithm of earnings to investments in human capital measured in time, such as years of schooling and years of post-school work experience. It has several desirable features:

(1) It is not an ad hoc specification. It is derived from an identity. As a result, the coefficients of the equation have economic interpretations.

---

* Research Professor, Department of Economics, University of Illinois at Chicago. This paper was written while Barry Chiswick was a Visiting Scholar in the Education Group, Human Development Department, World Bank. He appreciated the comments received on an earlier draft from Harry Patrinos, George Psacharopoulos and Jee-Peng Tan.

[1] The simple schooling version was first developed in Becker and Chiswick (1966), and extended to include on-the- job training in Mincer (1974).

(2) Because of the positive skewness of earnings and the rise in earnings inequality as schooling level increases, by using the natural logarithm of earnings rather than earnings as the dependent variable the residual variance in the HCEF is less heteroskedastic and the distribution of the residuals is closer to normal.

(3) It is an efficient user of data. Although data on earnings, years of schooling and years since leaving school are readily available, data on individual schooling costs are not readily available. The HCEF procedure involves converting a relationship between earnings and dollar investments in human capital to one between the natural logarithm of earnings and years of investment in schooling and training.

(4) The HCEF is flexible, allowing for easy incorporation of additional variables appropriate for the particular purpose of the study.

(5) Finally, the coefficients of the HCEF are devoid of units, facilitating comparisons across space (e.g., countries) or across time periods (e.g., decades).

One feature of the HCEF is its frequent use for estimating the rate of return from schooling. The coefficient of the schooling variable is often interpreted as the rate of return from schooling (see, for example, Psacharopoulos and Mattson 1996, Ram 1996, Rosen 1987, Willis 1986). While this may be the correct interpretation in some circumstances, it will be shown here that in principle, and in many circumstances, this is not the correct interpretation. This paper will also discuss the effects on the coefficient of schooling of the treatment of labor supply (weeks worked and hours worked per week) and other measures of labor market outcomes, such as occupational status.

II. Derivation of the Coefficient

For simplicity of presentation, what follows in this section will ignore post-school investments in on-the-job training and other variables, and will focus exclusively on the schooling variable. Let:

E_o = Earnings if there is no schooling ,

E_t = Earnings received each year after obtaining t years of schooling,

C_t = Dollar amount of investments in year t of schooling,

r_t = Rate of return on investments in year t of schooling,

$K_t = C_t \div E_{t-1}$ = Investment in level of schooling t relative to a full-year's potential earnings if investments were not made in this level of schooling.

If there is one period of investment in schooling for the individual, earnings after schooling is completed are:

(1) $$E_1 = E_0 + r_1C_1 = E_0 + r_1K_1E_0=E_0(1+r_1K_1)$$

For two periods:

(2) $$E_2 = E_1 + r_2C_2=E_1 + r_2K_2E_1 = E_1(1+r_2K_2)=E_0(1+r_1K_1)(1+r_2K_2)$$

Using the principle of mathematical induction,

(3) $$E_S = E_0 \prod_{t=1}^{S} (1+r_t K_t)$$

where S is the number of years of schooling completed. Taking natural logarithms,

(4) $$LnE_s = LnE_o + \sum_{t=1}^{s} Ln(1 + r_t K_t)$$

If $r_t K_t$ is small, we can apply the ruleregarding natural logarithms Ln (1 +ϵ) $\approx$ ϵ for small values of ϵ.[2] Then,

(5) $$LnE_s = LnE_o + \sum_{t=1}^{s} (r_t K_t)$$

Separate values of $r_t K_t$ can be estimated for each level of S, either individual years or groups of years (grade level). For simplicity of exposition, assume r_t and K_t do not vary with years of schooling ($r_o = r_t$ for all t and $K_o = K_t$ for all t). Then,

(6) $$LnE_s = LnE_o + (r_o K_o) S.$$

---

[2] For alternative values of Ln(1+ϵ):

Ln(1+ϵ)	ϵ
.0100	.01015
.0500	.0513
.1000	.1052
.1500	.1618
.2000	.2214
.3000	.3500
.4000	.4918
.8000	1.2255

The smaller is ϵ, the closer is the approximation Ln(1+ϵ) $\approx$ ϵ.

III. Interpreting the Coefficient

Adding a residual to equation (6) and estimating the regression equation, the coefficient of schooling is an estimate of the average percent increase in earnings per year of schooling.[3] Note that the coefficient of S is not the rate of return from schooling, but rather is rK. If the parameter K is known and the regression coefficient (b) is estimated, the rate of return from schooling is $r = b/K$.

The parameter K = 1 if the investment in schooling equals the full-year potential earnings if there were no further investment. This assumption was made to simplify the exposition in the orginal formulation of the specification (Becker and Chiswick, 1966). This assumption is also made in later treatments (see, for example, Mincer 1974, Willis 1986, Rosen 1987, Psacharopoulos and Mattson 1996). In most estimates of the rate of return from schooling using the HCEF there is no acknowledgment that this assumption is made; the coefficient of schooling is just accepted as the rate of return from schooling.

There are certain circumstances in which K=1. $C_t = E_{t-1}$ can occur, for example, if there are no out-of-pocket costs ($C_d = 0$) and the foregone earnings or opportunity cost ($C_f = E_{t-1}$) is a full year's earnings. It would also occur if, for example, opportunity costs were 75 percent of full-year potential earnings and it just so happened that direct costs were equivalent to 25 percent. However, K need not equal unity. Consider a case in which the direct costs of school are fully funded by the government, including books and school supplies. The student can work during school breaks, so the forgone earnings do not equal a full-year of potential earnings, but only 60

---

[3] Mincer (1974) shows that the coefficient of schooling (S) is biased downward if years of labor market experience is not included in the equation.

percent of a full-year. Then K = 0.60, and the rate of return is r = b/K = (1.67)b. A coefficient of schooling of .06 (or, six percent), which might seem low, would imply a rate of return of .10 or 10 percent.

Alternatively, consider a situation in which the student pays for tuition and all school fees and supplies, and other out-of-pocket expenditures related to schooling. Suppose the direct and foregone earnings equal 150 percent of potential earnings or K = 1.5. Then if b = .06, r = .06/1.5 = 0.04 or 4 percent. Very different interpretations emerge depending on the value of K.

The value of r or K for a country need not be constant across schooling levels. One may think of three levels of schooling, years of primary (P), years of middle or secondary (M), and years of tertiary or higher (H) education. Then,

(7) $$LnE_s = LnE_o + (r_p K_p)P + (r_m K_m)M + (r_h K_h)H,$$

where S = P+M+H.

The K's may vary by level of schooling. If, as it does in many developing countries, secondary schooling involves tuition charges as well as foregone earnings, K_m may exceed unity; if higher education involves subsidized tuition, fees and living expenses (i.e., lowering forgone earnings), K_h may be less than unity. Interpretations regarding the relative rates of return from different levels of schooling can be influenced by the estimated values of K.

An alternative procedure is to use dummy variables for each year of schooling or for each level of schooling. For example, we can define D_p, D_m and D_h as dummy variables that are unity if the person has completed primary (p), secondary (m), and higher (h) education, respectively.

Then the regression coefficient of the dummy variable D_j is $r_jK_jS_j$, where S_j is the number of years of schooling for education level D_j.

Thus far, the discussion has been in terms of private costs and private benefits. The implied rate of return is a private rate of return. The HCEF procedure allows for the computation of the "social" rate of return (r^*), defined as the rate of return based on total costs (private and social or public costs) and private benefits.[4] The regression coefficient is still b, but the interpretation is b = r*K*, where K^* incorporates the social cost of the investment.

Consider the following scenario: Foregone earnings constitute 75 percent of potential full-year earnings, and there are no tuition charges, school fees or other out-of-pocket expenses paid by the individual, but the cost to the public not paid by the individual is 50 percent of a student's full-year potential earnings. The regression coefficient is 0.06. The private rate of return is:

$r = b/K = .06/.75 = 0.080$, or 8.0 percent.

The social rate of return is:

$r^* = b^*/K^* = .06/1.25 = 0.048$, or 4.8 percent.

With an estimate of K and K*, which can be estimated on a group basis, not needing micro-level or individual data, both private and social rates of return can be computed.

Recall that if we do not make use of the approximation regarding the natural logarithm of a number close to one:

The private rate of return is, $r = \dfrac{(e^{b} - 1)}{K}$

---

[4] This concept of the social rate of return does not include what are referred to in the literature as the externalities from schooling.

Similarly, the social rate of return is, $r* = \frac{(e^{b}-1)}{K*}$

Hence the typical assumption that the coefficient of schooling in the HCEF is the rate of return from schooling is not correct in principle, and is approximately correct only if K ≈ 1 and the regression coefficient is a small number.

IV. Effects of other Explanatory Variables

The HCEF is highly adaptable, and a variety of variables have been added to the right-hand side. The most important and imaginative of these has been post-school labor market experience in Mincer's classic study (1974). Mincer shows that the coefficient of schooling is biased downward if age rather than labor market experience is held constant.[5] Moreover, he develops the rationale for the now standard quadratic form for the experience variable.

Measures of employment (labor supply) can be incorporated into the analysis. Let E_a equal annual earnings and E_h equal hourly earnings, where W is weeks worked in the year and H is

---

[5] If $LnY = b_o + b_1 S + b_2 T$, where T equals Age-S-Z, that is, potential labor market experience is the number of years since leaving school (S+Z), where Z is the age at which school is started. Then $LnY = b_o + b_1 S + b_2 (A\text{-}S\text{-}Z) = (b_o - Zb_2) + (b_1 - b_2) S + b_2 A$. In some circumstances, especially in developing countries where schooling attainment is very low, potential labor market experience (T) should be measured a non-negative number which is as the lesser of:

T = Age-S-Z, and
T = Age - X,

where X is the age of the onset of employment that provides labor market experience relevant for adult work. The level of potential labor market experience may also need to be adjusted for grade repetition.

hours worked per week. Then, $E_a = E_h(W)^{\gamma_1}(H)^{\gamma_2}$, where γ_1 i = 1,2, is the elasticity of

earnings with respect to time worked. Then,

(8) $LnE_a = LnE_h + \gamma_1(LnW) + \gamma_2(LnH),$

where LnE_h is replaced by the human capital earnings function variables, schooling, experience, etc. Suppose the elasticity of earnings with respect to time units worked is unity, that is, γ_1 and γ_2 are unity. Then, a 10 percent increase in time units worked (weeks or hours) increases annual earnings by 10 percent.

Note, however, that labor supply decisions (time worked) are not exogenous with respect to wage rates. This raises issues of endogenous explanatory variables which are typically ignored. As a result, however, the elasticity of earnings with respect to time units worked may exceed unity. This arises if labor supply curves are upward rising. It also arises if data are available for weeks worked but not hours worked per week, and yet weeks worked and hours worked per week are positively correlated.

Alternatively, the elasticity of earnings with respect to time units worked may be less than unity. This arises if labor supply curves are backward bending (i.e., the income effect of a higher wage dominates the substitution effect) or if there is anticipated variable period (peak-load) employment. Seasonality of employment, for example, would result in higher weekly earnings, but lower annual earnings for those who work fewer weeks in the year if, say, there is unemployment in the off season which is compensated at a lower rate. A similar situation may arise for hours worked per week. Purely random measurement errors will tend to bias γ downward, resulting in a measured γ less than unity even if the true population value is unity.

The expression in equation 8 permits the estimation of the elasticities of earnings with respect to time units worked, where the interesting null hypothesis is that $\gamma_1 = \gamma_2 = 1$. Empirically, for developed countries the elasticity of earnings with respect to weeks worked is often close to unity, while the elasticity with respect to hours worked is substantially and significantly below unity.

This issue is particularly relevant because of a tendency to convert the dependent variable from annual earnings into hourly earnings. Then, transforming equation (8),

$$(9) \quad Ln\left(\frac{E_a}{WH}\right) = LnE_h + (\gamma_1 - 1)Ln(W) + (\gamma_2 - 1)Ln(H).$$

However, when hourly earnings is the dependent variable, whether observed or constructed, variables for weeks worked per year and hours worked per week are generally not included in the regression equation. Then, to the extent that γ_i does not equal unity and the time worked variable is correlated with an included variable, the coefficient of the included variable is biased. For example, suppose $\gamma_1 < 1$ and $\gamma_2 < 1$ and that those with more schooling work more weeks in the year and more hours per week. If the dependent variable is computed hourly earnings (E_a/WH), the coefficient of schooling is biased downwards. This results in a downward biased estimate of the rate of return from schooling. Another variable that can bias the coefficient of schooling is an alternative measure of labor market success. Consider the effects of putting occupational dummy variables, an occupational prestige score, or a variable for living in a low-income area on the right hand side of the equation. Then the regression estimates the effect of schooling on earnings within an occupational or household income strata. The coefficient does not incorporate the effect of schooling in raising occupational level and household income. It is not that this approach is "wrong" but that it tends to result in incorrect interpretations. Consider the following equation:

(10) $\boldsymbol{LnE = b_0^1 + b_1^1 S + b_2^1 OCC}$,

where OCC = 1 for professionals and managers and OCC = 0 otherwise. The effect of schooling on earnings is:

$$\frac{\partial LnE}{\partial S} = b_1^1 + b_2^1 \frac{\partial OCC}{\partial S}. \tag{11}$$

The coefficient of schooling b_1^1 is a downward biased estimate of the partial effect of schooling on earnings. It merely measures the average effect of schooling on earnings within occupational strata. The coefficient b_2^1 is the effect of a higher occupational strata on earnings and $\frac{\partial OCC}{\partial S}$ is the effect of schooling on occupational status. It is not the coefficient of schooling, but rather $\frac{\partial LnE}{\partial S} = b_1^1 + b_2^1 \frac{\partial OCC}{\partial S}$ that is the estimated value of rK. Similarly, stratifying the data by measures of labor market success (e.g., within occupational categories) and computing regression equations within strata also biases downward the coefficient of schooling.

When interpreted correctly, the HCEF can be an invaluable tool for estimating the rate of return from schooling.

References

Becker, Gary S. and Barry R. Chiswick (1966). "Education and the Distribution of Earnings."American Economic Review, Vol. 56, May, pp. 358-69.

Federal Judiciary Center (1994). Reference Manual on Scientific Evidence. Washington, D.C.

Gastwirth, Joseph (1988). Statistical Reasoning in Law and Public Policy. Academic Press, San Diego.

Mincer, Jacob (1974). Schooling, Experience and Earnings, New York: National Bureau of Economic Research.

Psacharopoulos, George and Robert Mattson (1996). "Estimating the Returns to Education: A Sensitivity Analysis of Methods and Sample Size." World Bank, Xerox.

Ram, Rati (1996). "Level of Development and Rates of Return to Schooling: Some Estimates from Multicountry Data." Economic Development and Cultural Change, Vol. 44, No. 4, July, pp. 839-857.

Rosen, Sherwin (1987). "Human Capital." in John Eatwell, et al., eds., The New Palgrave: A Dictionary of Economics, London: Macmillian, pp. 681-90.

Willis, Robert J. (1986). "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions." in Orley Ashenfelter and Richard Layard, eds. Handbook of Labor Economics, Vol. I, pp. 525-601.