한국경제학보 제13권 제1호
The Korean Journal of Economics Vol. 13, No. 1(Spring 2006)

The Effect of Omitted Variables in Education and Earnings

Sung-Shin Han* · Insook Cho**

Abstracts

Many studies on human capital and earnings have reported a relative 5~15 percent wage difference per additional year of schooling. It is difficult to interpret such earnings gaps as a credible estimate of the economic return from more schooling without understanding the role of one's ability in the education process. This paper attempts to illustrate the potential role of unobservable ability and show evidence on the omitted variable bias, using control variables for ability and the IV estimation method. The results provide some support in favor of the hypothesis that the conventional OLS estimates are upwardly biased due to the unobservable ability variable in the earnings equation.

KRF Classification : B030400
Keywords : return to education, human capital, unobservable ability, omitted variable bias

Ⅰ. Introduction

Workers who attained a higher level of education are more

* First author, Department of Economics, Yonsei University, Seoul, Korea (e-mail: sshan@yonsei.ac.kr)
** Second author, Department of Economics, University of California, Santa Barbara, California, U.S.A. (e-mail: icho@econ.ucsb.edu)
Authors would like to thank to anonymous referees for comments on an earlier draft. All remaining errors are all our own.

likely to receive higher wages. Hundreds of studies on human capital and earnings have reported a relative 5~15 percent wage difference per additional year of schooling. Unless we can observe the earnings of identical individuals with more or less education, it is difficult to interpret such earnings gaps as a credible estimate of the economic return from more schooling. An individual may receive higher wages because he/she accumulates higher level of human capital in school. On the other hand, an individual with greater earning capacity may choose staying school longer.

Higher level of human capital accumulation became one of the important factors that explain productivity growth. Becker's (1964) landmark book stimulated many labor economists to study the role of education on labor market outcome. Many people were skeptical about the interpretation of the economic returns from schooling. Denison (1964) claimed that we could attribute different education levels (of workers) to wage differentials (between workers with different education levels) only if people attain different levels of education regardless of their inherent ability level. Grilliches (1977) explained that it might be true the estimated return to schooling were upwardly biased because more capable people were more likely to attain higher level of education. He tried to convince that such an upward bias can be offset with measurement error bias. Recent studies about the role of education in labor market outcomes have approached more carefully to this omitted variable problem. Altonji and Dunn (1996) showed that it is possible to obtain less biased estimates by adding more control variables in Mincer (1974)- type earning equation. Angrist and Kruger (1991), Card (1995) and Maluccio (1997) tried to find unbiased return to schooling using various instrumental variables. Studies using a with-in family pair data received much attention from people.

Understanding the role of one's ability in the education process is a critical component of most public policy discussions about the effectiveness of investment in education. Attempts to understand the influence of schooling on the increasing income inequalities observed in recent years have heightened the concern with these issues.

This paper illustrates the relationship between education and earning using a simpler utility maximization model. The model focuses on the potential role of unobservable ability. The genuine role of education on future income can be assessed by estimating the "true" return to education.

This paper consists of five sections. Section Ⅱ describes the data set used in this paper. Section Ⅲ illustrates a simple optimization model of schooling and discussed issues related to omitted variables in an estimation of the return to education. Section Ⅳ presents the interpretations of empirical results. Concluding remarks are in Section Ⅴ.

Ⅱ. The data

The data set used in the analysis is The National Education Longitudinal Study of 1988 (NELS: 88). The NELS is a nationally representative sample of U.S. students at their 8^{th} grade during the spring semester of 1988.[1] After the initial survey in 1988, four more follow-up interviews have been conducted in 1990, 1992, 1994 and 2000. Only 12,144 members remained in the sample in

---

1) In Korea, we can find an on-going project constructing a similar type of data base to the NELS: 88. The Korean Education and Employment Panel (KEEP) surveys students at their 9^{th} and 12^{th} grade during the school year 2004 in Korea. The survey will follow up these students' future labor market outcomes for the next 10 years of the first survey.

the forth follow-up study of 2000.[2] The NELS data set provides an extensive set of information about students' secondary and postsecondary education, labor market outcomes, community integration, family background and marriage and family formation. Sample mean and standard deviations for key variables are presented in Table 1.

Ⅲ. Methodology

1. Basic optimizing model of school choice

In order to understand the role of unobservable ability, we set up a simple optimizing model of schooling choice. This model follows the standard human capital earnings function which states the log earnings is a linear function of education (Mincer, 1974.)

We assume that individuals seek to maximize utility, which is a function of income and schooling:

$$U(y, S) = \ln(y) - f(S) = \ln[g(S)] - f(S) \tag{1}$$

where $y = g(s)$ represents the observable relationship of earnings (y) to schooling (S), and $\ln[g(s)]$ and $f(s)$ are increasing convex functions that represent the (log) benefits and costs of schooling. Maximizing utility in equation (1) requires that optimal schooling

2) Originally NELS: 88 dataset was designed to collect data from the nationally representative samples in the U.S. Through four times of follow-up studies, some samples did not respond or could not be reached. After the fourth follow-up study, only 12,144 samples remained. Weights are used to compensate for unequal probabilities of selection and to adjust for the effects of non-response. While the weight generalizes to no meaningful analysis population, when used in conjunction with appropriate population definitions, it can be used to estimate parameters that describe the populations of spring 1988 8th-graders in the year, 2000.

(S^*) satisfy the first-order condition,

$$g'(S)/g(S) = f'(S), \tag{2}$$

where marginal benefits ($g'(S)/g(S)$) are equal to marginal costs ($f'(S)$).

The choice of functional form for the marginal benefits and costs of schooling is important for an empirical implementation of this model. To capture the well-known stylized fact that log earnings is a nearly linear function of schooling, it must be the case that for individual i, the marginal benefit (MB_i) of schooling is represented by,

$$MB_i = g'(S)/g(S) = b_i + \theta A_i, \tag{3}$$

where A_i is unobserved "ability" of the individual. To generate an interior solution for the choice of schooling, we assume that the MC_i of schooling has the simple form,

$$MC_i = f'(S) = r_i + r_0 S_i. \tag{4}$$

It follows that the optimal level of schooling is

$$S_i^* = \frac{b_i - r_i}{r_0} + \frac{\theta}{r_0} A_i, \tag{5}$$

which varies across individuals according to the difference of the marginal rate of return (b_i) and the marginal cost for schooling (r_i) for each individual i. From the equation (5), it is clear that an individual's schooling choice is affected by the unobserved "ability" of the individual if θ differs from zero.

2. The average return to education

A standard earnings equation shows us the relationship between income and schooling. An individual's earning depends on schooling (S_i), the individual's ability (A_i), other observable characteristics (X_i), and an error term ϵ_i.

$$y_i = A_i + b_i S_i + dX_i + \epsilon_i. \tag{6}$$

where y_i, S_i, X_i, and A_i denote the logarithm of the wage rates, the schooling level, other observable characteristics and an unobserved ability. From equation (6), we can estimate the average return of education $\bar{b}$.[3]

$$plim\, b_{ols} = cov(y_i, S_i)/var(S_i). \tag{7}$$

Since the ability variable A_i is unobservable, the ordinary least squares estimator (b_{ols}) in equation (7) may suffer from the omitted variable bias. Equation (8) shows the size of bias caused by the unobserved ability term A_i.

$$plim\, b_{ols} = \bar{b} + cov(A_i, S_i)/var(S_i). \tag{8}$$

3) This model assumes a heterogeneity for the return to schooling (b_i) for each individual i. The assumption may complicate the analysis for the causal effect of education on earnings. Estimation of returns to education using earning equations in this paper calculates an average return across the entire population, $\bar{b}$, assuming individual's return to schooling (b_i) is uncorrelated with an error term (ϵ_i). This interpretation fits the goal of the analysis in this paper.

The conventional OLS estimation for the return to education is biased because an individual's schooling choice is correlated with one's ability level. To correct this omitted variable bias, three alternative econometric methods can be used. The first is to add more control variables which capture the impact of the unobserved ability (A_i) on earnings (y_i). The second solution is the two-stage least squares estimation (TSLS) using instrumental variables. The last solution is to construct a fixed effect model using a panel data set. The following sub-sections illustrate three methodologies.

3. Variables for ability control.

One solution to fix the omitted variable bias is to add more control variables which can capture the impact of the unobserved ability in the earnings function. Suppose there is an observable variable IQ_i which can represent an individual's ability level. The relationship between this new variable (IQ_i) and ability can be expressed as the following.

$$IQ_i = \eta A_i + v_i. \tag{9}$$

where ν_i is a random error. Substituting equation (9) for unobservable ability variable in equation (6), we can rewrite the earnings function (6) as a function of education (S_i) and the new variable (IQ_i).

$$y_i = IQ_i + b_i S_i + d' X_i + \epsilon_i'. \tag{10}$$

$$\begin{aligned} plim\, b_{ols} &= cov(y_i, S_i)/var(S_i) \\ &= b + cov(IQ_i, S_i)/var(S_i) \end{aligned} \tag{11}$$

If the new variable IQ_i is not correlated with the schooling level S_i, then this new earnings equation provides unbiased estimates for the return to education.

It is difficult to find a perfect proxy that is uncorrelated with one's schooling choice and correlated with earnings. However, one advantage of using this method is that it is possible to estimate less biased return to schooling.

4. Instrumental variables estimation

Another solution to fix this endogeneity problem is to use the two-stage least squares estimation method. Suppose there is an instrumental variable Z_i which satisfies the following two criteria: i) $cov(S_i, Z_i) \neq 0$, ii) $cov(A_i, Z_i) = 0$ and $cov(\epsilon_i, Z_i) = 0$. The instrumental variable, Z_i, allows to implement the following two-stage least squares estimation model.

$$S_i = a_0 + a_1 Z_i + e_i \tag{12}$$

$$y_i = A_i + b_i \hat{S}_i + d'' X_i + \epsilon_i''. \tag{13}$$

$$\begin{aligned} plim\ b_{TSLS} &= cov(y_i, \hat{S}_i)/var(\hat{S}_i) \\ &= b_i + cov(A_i, \hat{S}_i)/var(\hat{S}_i) \\ &= b_i + cov(A_i, Z_i)/var(Z_i) \\ &= b_i \end{aligned} \tag{14}$$

Education variable fitted by an instrumental variable in the first stage equation (12) substitutes a real education variable in the second stage regression equation (13). The TSLS estimates from this equation provide unbiased estimates for the return to education as equation (14).

Theoretically the instrumental variable method is a promising way to estimate a "true" economic return to education. However, it is difficult to find a good instrumental variable for an individual's schooling decision. Many economists have attempted to use many instrumental variables, including birth cohorts, proximity to colleges, and compulsory education law.

5. The fixed effect model using panel data set.

An alterative way to correct an omitted variable bias is to study education and earnings outcomes for siblings, twins, or father-son/mother-daughter pairs. The key idea behind this strategy is that some of the unobserved differences that can create bias in a cross-sectional comparison of education and earnings are reduced or eliminated within families. The two earnings equations for person 1 and person 2 from the same family j provide an illustration of how the ability effects are eliminated within families.

$$y_{1j} = A_j + b_{1j}S_{1j} + dX_{1j} + \epsilon_{1j},$$
$$y_{2j} = A_j + b_{2j}S_{2j} + dX_{2j} + \epsilon_{2j}. \qquad (13)$$

y_{1j} and y_{2j} are incomes, S_{1j} and S_{2j} are the schooling decisions, X_{1j} and X_{2j} are other observable characteristics, and ϵ_{1j} and ϵ_{2j} are random errors of person 1 and person 2 from family j. The key assumption for using these earning equations of two family members is that the ability levels of two persons from the same family $j(A_j)$ are the same and the return to education is the same, i.e., $b_{1j} = b_{2j} = b_j$. Under that assumption, a fixed effects model can be constructed by differencing out two equations. Equation (16) is the fixed effect earning equation:

$$(y_{1j} - y_{2j}) = \bar{b}(S_{1j} - S_{2j}) + d(X_{1j} - X_{2j}) + (\epsilon_{1j} - \epsilon_{2j}) \qquad (16)$$

Under the assumption that the ability levels of two persons 1 and 2 from the family j are the same, unobserved ability variable can be eliminated from our earnings equations. The estimated return to education using equation (16) provides unbiased estimates for the return to education.

$$plimb_{ols} = cov((y_{1j} - y_{2j}), (S_{1j} - S_{2j})) / var((S_{1j} - S_{2j})) \\ = \bar{b} \tag{17}$$

This method also has shortfalls. First, there is no systematic evidence that the ability levels of any two members of a family are the same, even the case of twins. Second, the results from estimations using within family differences should be interpreted with caution. For example, researchers may be reluctant to say that the entire population can be well-represented by a sample of twins. Third, within-family analysis may limit the number of control variables to use. For example, any two members of a family are more likely to live in the same state; have the same family background; and have the same racial/ethnic background. It the impacts of those variables were not captured, the estimates may suffer from bias.

Griliches (1979) showed that the key question is whether the fixed effects estimator leads to less bias in the estimated return to schooling than the simple cross section regress estimator. He suggested that the bias in the fixed effects estimator will be smaller than the bias in the cross-section estimator if the regression coefficient of ability on schooling is lower in the within-twins regression than in the across twins regression.

Ⅳ. Results

1. The average return to education

Table 2 reports the average return to education estimated using equation (6). The coefficient for the college dummy is 0.29, which implies the wage gap between college graduates and high school graduates is 29 percentage points. The coefficient for years of education is 0.9. This estimate implies that if an individual attains one more year of education, then his/her future earnings are higher than those of lower educated counterparts by 9 percentage points. Table 2 also illustrates other stylized facts, such as the positive effect of work experiences on earnings, wage differentials between genders, the effects of family formation and earnings difference among different regions.

This simple regression of wages on education levels shows that an individual with higher education is more likely to earn higher wages. Equation (8) suggests that these estimates may suffer from the omitted variable bias. New estimates using three different econometric methodologies are illustrated in the following sections.

2. Control for unobservable ability.

Test scores[4)] of students and parents' education as potential indicators of ability. From the estimation of the standard earnings function as (6), the impact of ability on earnings is usually not captured. The estimates in this section suggest that the

4) The students' grades used in the estimation are the combinations of the grades from four different subjects (math, English, history and science) at their 8^{th}, 10^{th} and 12^{th} grade levels. The GPA distribution at colleges is also used to capture the impact of ability. Generally these grades are considered to represent a person's ability relatively well.

[Table 1] Description of the data

	Overall sample		Male	Female
	Mean	Std.Dev.	Mean	Mean
Number of observations	11,798		5,677	6,121
Education				
Years of completed education	14.29 years	1.92	14.15 years	14.56 years
Dropout	5.49%	0.22	5.78%	5.21%
High school graduates	20.45%	0.40	23.00%	18.09%
Some college experiences	44.73%	0.49	45.02%	44.47%
B.A.	29.91%	0.45	27.97%	31.71%
M.A.	3.30%	0.17	2.48%	4.05%
Ph.D.	0.65%	0.08	0.65%	0.65%
Earnings and working conditions				
Employment status	86.86%	0.34	91.54%	82.71%
Hourly wage	14.95 $/hr	14.27	16.51 $/hr	13.49 $/hr
Income in 1999	23,379 $/yr	20,563	23,653 $/yr	19,003 $/yr
Working hour per week	39.84 hr/wk	12.09	42.84 hr/wk	37.06 hr/wk
Working week per year	42.05 wk/yr	17.57	44.71 wk/yr	39.59 wk/yr
Wage growth (97-99)	0.47	2.76	0.49	0.45
Experience	3.24 years	2.29	2.39 years	3.11 years
Personal characteristics				
Male	48.11%	0.49		
White	68.75%	0.46	69.58%	67.99%
Black	9.60%	0.29	9.00%	10.16%
Hispanic	13.23%	0.34	12.84%	13.59%
Asian	7.06%	0.25	7.29%	6.86%
Marrital status	39.48%	0.48	34.65%	43.98%
Number of kids	0.62 kid	0.97	0.51 kids	0.73 kids
Urban area	28.06%	0.44	27.97%	28.14%
North east	18.76%	0.39	18.51%	19.00%
North central	26.42%	0.44	26.63%	26.23%
South	33.16%	0.47	33.50%	32.85%
West	20.03%	0.40	19.76%	20.29%
Family background				
Father's education	13.97 years	3.51	14.08 years	13.87 years
Mother's education	13.38 years	2.81	13.58 years	13.29 years

[Table 2] The OLS regression for return to education

	(1)	(2)	(3)	(4)	(5)	(6)
	Coeff. (Std. Err)	Coeff. (Std. Err)	Coeff. (Std. Err)	Coeff. (Std. Err)	Coeff. (Std. Err)	Coeff. (Std. Err)
B.A. or higher	0.31** (0.01		0.29** (0.01)		0.29** (0.01)	
Years of education		0.10** (0.02)		0.09** (0.02)		0.09** (0.02)
Experience	0.05** (0.01)	0.05** (0.01)	0.04** (0.01)	0.04** (0.01)	0.04** (0.01)	0.04** (0.01)
Exp squared	-0.00** (0.00)	-0.00** (0.00)	-0.00** (0.00)	-0.00** (0.00)	-0.00** (0.00)	-0.00** (0.00)
Male	0.18** (0.01)	0.19** (0.01)	0.14** (0.01)	0.15** (0.01)	0.15** (0.01)	0.15** (0.01)
Black	-0.12** (0.01)	-0.12** (0.01)	-0.09** (0.14)	-0.10** (0.01)	-0.07** (0.01)	-0.08** (0.01)
Married			0.02 (0.01)	0.01 (0.01)	0.03* (0.01)	0.02 (0.01)
Male*married			0.09** (0.02)	0.09** (0.02)	0.09** (0.02)	0.09** (0.02)
No kid			0.08** (0.01)	0.07** (0.01)	0.08** (0.01)	0.07** (0.01)
Urban					0.03** (0.01)	0.03** (0.01)
North East					0.09** (0.01)	0.09** (0.01)
North Central					0.06** (0.01)	0.05** (0.01)
West					0.09** (0.01)	0.08** (0.01)
observation	10,635	10,558	10,635	10,558	10,635	10,558
R-squared	0.141	0.138	0.138	0.143	0.145	0.150

Note: All models are population weighted and heteroskedastic-consistent standard errors are in parenthesis. Coefficients with (**) are significant at the 2.5 percent level.

conventional OLS estimates are upwardly biased because of omission of ability.

Table 3 reports the estimated return to education with the test scores and parents' education variables. The estimates are slightly lower than the OLS estimates. That result is consistent with the assumption that the conventional OLS estimates are upwardly biased because of the omitted variable problem.

The estimates in Table 3 may have bias. First, test scores or parents' education information does not completely capture the impact of ability on earnings. Second, such control variables can be correlated with the schooling decision of the students. In particular, parent's education levels are strongly correlated with children's education levels.

[Table 3] OLS estimation controlling ability variable: Using grades and parental education.

	Schooling variable		Control variable		R squared
Control Variable	Coeff.	Std.Err	Coeff.	Std. Err	
OLS regression using a dummy variable for college graduates					
Father's education	0.28**	(0.01)	0.03**	(0.01)	0.150
Mother's education	0.28**	(0.01)	0.04**	(0.01)	0.150
Test score at the 12^{th} grade	0.28**	(0.01)	0.00	(0.00)	0.146
Test score at the 10^{th} grade	0.27**	(0.01)	0.001**	(0.00)	0.145
Test score at 8^{th} grade	0.27**	(0.01)	0.002**	(0.00)	0.147
GPA at college	0.25**	(0.01)	0.015**	(0.00)	0.150
OLS regression using year-measurement for education					
Father's education	0.08**	(0.02)	0.03**	(0.01)	0.150
Mother's education	0.08**	(0.02)	0.04**	(0.01)	0.150
Test score at the 12^{th} grade	0.09**	(0.02)	-0.00	(0.00)	0.150
Test score at the 10^{th} grade	0.08**	(0.02)	0.001**	(0.00)	0.151
Test score at 8^{th} grade	0.08**	(0.02)	0.001**	(0.00)	0.152
GPA at college	0.08**	(0.02)	0.00	(0.00)	0.150

Note: All models are population weighted and standard errors are heteroskedastic consistent. Coefficients with (**) are significant at the 2.5 percent level.

3. IV estimation

Family background information is often considered as promising instrumental variables for unobserved ability in many estimations of earning functions. However, economists are cautious of using such information as legitimate instrumental variables. For example, parents' education levels might be a good proxy for students' unobserved ability. At the same time, we know that there exists a strong correlation between parents' education level and their kid's education level. This correlation violates one of the important requirements as an instrumental variable and it can even bring about a larger bias in estimates for the returns from schooling.

[Table 4] TSLS estimation for the return to education using birth cohort variable

	OLS	TSLS	OLS	TSLS
B.A. or higher	0.30**	0.77**	0.30**	0.58**
	(0.01)	(0.14)	(0.01)	(0.12)
Years of education			0.08**	0.13**
			(0.0027)	(0.0480)
Male	0.21**	0.23**	0.21**	0.23**
	(0.01)	(0.01)	(0.01)	(0.01)
Black	-0.08**	-0.01	-0.09**	-0.07**
	(0.01)	(0.02)	(0.01)	(0.03)
Married	0.04**	0.08**	0.03**	0.01
	(0.01)	(0.02)	(0.01)	(0.14)
Urban	0.03**	-0.01	0.03**	0.06**
	(0.01)	(0.02)	(0.01)	(0.02)
Year dummy	Yes	Yes	Yes	Yes
Region dummy	Yes	Yes	Yes	Yes
R squared	0.14	0.09	0.15	0.11

Note: All models are population weighted and heteroskedastic-consistent standard errors are in parenthesis. Coefficients with (**) are significant at the 2.5 percent level.

In recent literature, much attention has focused on institutional sources of variation in schooling. These attribute to such features as the minimum schooling leaving age (Angrist and Krueger, 1991), tuition costs for higher education, (Kane and Rouse, 1995) or the geographical proximity of schools (Card, 1995). Such institutional factors have a reasonable chance of satisfying the strict exogeneity assumptions required for a legitimate instrumental variable. The Angrist and Krueger (1991) paper is the landmark in such trend of literatures. They established that season of birth is related to educational attainment because of school starting age policy and compulsory school attendance laws. Individuals born at the beginning of the year start school at an older age and can therefore drop out after completing less schooling than individuals born near the end of the year. Angrist and Krueger estimate the impact of compulsory schooling on earnings by using a quarter of birth as an instrument for education. They report that the instrumental variables estimate of the return to education is close to the ordinary least squares estimate, suggesting that there is little bias with conventional estimates.

In this paper, the TSLS estimates are calculated using compulsory school law as an instrumental variable.[5] Table 4 shows that the TSLS estimates are higher than OLS estimates. From equation (12), the TSLS estimates are expected to be lower than the OLS estimates. One explanation is a measurement error problem. Suppose some of our samples reported their completed education levels higher than true ones.

$$S_i = S_i^{true} + \xi_i . \tag{18}$$

S_i is observed education information, S_i^{ture} is the true education

5) Compulsory school law affects only the people who want to drop out from high school before some legal age. It does not affect people's decision about the attendance to postsecondary school. Since the dummy variable I use in the estimation is for the people who graduate college, the birth cohort is not a meaningful instrumental variable for this education variable. Conclusively, if we use this birth cohort as an instrumental variable for this dummy variable for education, it becomes a difficult task to interpret the coefficient for this instrumental variable.

level and ξ_i is the errors in self reported education level. OLS estimate with measurement error may suffer two different biases, assuming that this measurement error is not correlated with the ability term. In equation (19), the first term, the "true" return to education, is downwardly biased due to the measurement error problem. The second term, the bias caused by the omitted variable problem, is smaller in magnitude because of the measurement error. If the data has a measurement error problem, it is not clear whether OLS estimates using standard earnings equation would be upwardly biased or downwardly biased.

$$plimb_{ols} = \bar{b}\frac{var(S_i^{true})}{var(S_i^{true}) + var(\xi_i)} + \frac{cov(A_i, S_i^{true})}{var(S_i^{true}) + var(\xi_i)} \quad (19)$$

Research over the past three decades has generally found that the reliability of self reported schooling is about 90% in the United States. It suggests that both terms in the equation (19) appear smaller by 0.1 in most data sets. This implies that the OLS estimates in this paper might suffer from two different biases, the measurement error bias and the omitted variable bias. Considering that a measurement error problem can be avoided by using instrumental variable methods, the slightly higher estimates from IV estimation would be closer to the true value.

The second explanation is the local treatment effect. Angrist and Krueger stated in their paper that around 25 percent of high school students decide to stay at school longer because of compulsory school law. Conversely only 25 percent of the samples are affected by the compulsory school law and attend school longer than they want. Using the birth cohort variable as an instrumental variable for education level allows us to estimate the estimated return to education for those who are affected by the law, not the average return to education of whole sample. The estimated returns to education for those students, whose schooling decisions are affected by the existence of compulsory school law, are somewhat higher than the average marginal return to education of the entire population. In that case, the TSLS estimates may not be the average return to education over the entire population.

The estimates using the birth cohort variable as an instrumental variable are slightly higher estimate than the OLS estimates. If the date of birth of an individual is correlated with one's schooling decision, the estimates from this model would be the true average return to education.

4. The fixed effect model using the panel data set

Ashenfelter and Rouse (1998) developed a model of optimal schooling investments. They demonstrated the return to schooling using new data of approximately 700 identical twins. They reported that the estimated return to schooling was around 9 percentage points for the unique sample of identical twins. Emphasis was put on the potential role of unobservable ability in the determination of both schooling and income. The key assumption is that the schooling investment of identical twins should be the same, apart from random deviations that are not related to the determinants of schooling choices. It follows that contrasts of the earnings differences of identical twins with their education differences may provide a particularly useful way to isolate the causal effect of schooling on earnings. Researchers found that estimates of the return to schooling based on comparisons between brothers of fraternal twins contain some positive bias but less than the corresponding OLS estimates.

A similar type of analysis as Ashenfelter and Rouse had performed is not available using the NELS data set because the NELS data set does not provide information of paired individuals.

V. Conclusion

In this paper, we have explored the possibility that the marginal rates of returns from education are influenced by "unobserved" ability in the wage equation. A standard earnings equation cannot capture the impacts of an individual's ability on earnings and schooling. This paper provides empirical evidence on the omitted variable bias and suggests three different econometric methodologies that may eliminated the bias. Our results provide some support in favor of the hypothesis that the conventional OLS estimates are upwardly biased due to the unobservable ability variable in the earnings equation.

Consistent with much of the recent literature, the estimates using different econometric methods do not differ much from the conventional OLS estimates. The "best available" evidence from the latest studies of identical twins suggests a small upward bias in the simple OLS estimates.

Much attention has been paid to explain the recent increase of income inequality in the Korea. Even though many studies have reported positive relationship between education on earnings, there is no strong evidence that education has causal effects on future earnings regardless of one's inherent ability. Further studies may be required to show the true relationship among ability, education, and future earnings. Better understanding over the role of one's ability on education and earnings may help to answer the effectiveness of investment in education.

◈ *References* ◈

Altonji, Joseph G. and Thomas A. Dunn (1996), "The Effects of Family Characteristics on the Return to Education," *The Review of Economics and Statistics*, Vol. 78, No. 4, pp. 692-704.

Angrist, Joshua D. and Alan B. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, Vol. 106, No. 4, pp. 979-1014.

Ashenfelter Orley and Cecilica Rouse (1998), "Income, Schooling, and Ability: Evidence from a New Sample of Identical Twins," *The Quarterly Journal of Economics*, 113, pp. 253-284.

Blackburn, McKinley and David Neumark (1993), "Omitted-Ability Bias and the Increase in the Return to Schooling," *Journal of Labor Economics*, Vol. 11, No. 3, pp. 521-544.

Becker, Gary S. (1964), *Human capital: a theoretical and empirical analysis, with special reference to education*, New York: Columbia University Press.

Bound, John, David A. Jaeger and Regina M. Baker (1995), "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variables is Weak," *Journal of the American Statistical Association*, Vol. 90, pp. 443-450.

Card, David (2001), "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, Vol. 69, No. 5, pp. 1127-1160.

__________, (1999), "Causal Effect of Education on Earnings," in Orley Ashenfelter and David Card, editors, *Handbook of Labor Economics*, Volume 3A (North Holland, Amsterdam and New York).

__________, (1995), "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," *Aspects of Labour Economics: Essays in Honour of John Vanderkamp, Edited by Louis Christofides*, E. Kenneth Grant and Robert Swindinksy

(University of Toronto Press).

Curtin, T.R., Ingels, S.J., Wu, S., and Heuer, R. (2002), *National Education Longitudinal Study of 1988: Base-Year to Fourth Follow-up Data File User's Manual (NCES 2002-323)*, Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Denison, Edward F. (1964), "Measuring the contribution of education," in *The residual factor and economic growth*, Paris: OECD.

Griliches, Zvi (1979), "Issues in Assessing the Contribution of Research and Development to Productivity Growth," *Bell Journal of Economics*, The RAND Corporation, Vol. 10, No. 1, pp. 92-116.

__________, (1977), "Estimating the Returns to Schooling: Some Econometrics Problems," *Econometrica*, Vol. 45, pp. 1-22.

Hungerford, Thomas J. and Gary Solon (1987), "Sheepskin Effects in the Return to Education," Review of Economics and Statistics, Vol. 69, No. 1, pp. 175-177.

Kane, Thomas J. and Rouse, Cecilia Elena (1995), "Labor-Market Returns to Two- and Four-Year College," *American Economic Review*, Vol. 85, No. 3, pp. 600-614.

Maluccio, John (1997), "Endogeneity of Schooling in the Wage Function," Unpublished Manuscript, Yale University Department of Economics.

Mincer, Jacob (1974), *Schooling, Experience, and Earnings* New York: Columbia University Press.

Park, Jin Heum (1994), "Return to Schooling: A Peculiar Deviation from Linearity," Princeton, Department of Economics, Industrial Relation Section Working Paper, No. 335.

개인 능력변수의 누락이 교육투자수익률의 추정에 미치는 영향에 대한 분석

한 성 신 · 조 인 숙

논문초록

그간 많은 연구를 통해, 교육투자의 수익률은 약 5-15 % 정도일것이라 추정되고 있다. 각 개인의 능력과 그의 교육 및 임금에 미치는 영향을 고려해본다면, 기존의 교육투자수익률 추정치를 순수히 교육을 통한 임금수준의 격차로 해석하기는 무리가 따른다. 이 논문은 객관적 측정이 어려운 개인별 능력이 교육투자수익률 추정에 어떤 영향을 미치는지를 설명하고, 그에 대한 실질분석 자료를 제시한다. 이 논문에서는 전통적인 OLS방식을 이용한 교육투자수익률 추정은 실제 수익률을 과대 추정하고 있을 가능성을 제시한다.

주제분류 : B030400
핵심 주제어 : 관찰불능 능력변수, 교육수익율, 도구변수추정, 변수의 측정오류