The Private Return on Education and How to Solve the Endogeneity Problem:  
Case Indonesia

Nenny Hendajany  
Department Economics  
Universitas Sangga Buana  
Bandung, Indonesia  
neni.hendajany@usbypkp.ac.id

Abstract—This paper discusses the return on education using the Mincer model. The Mincer equation is a log natural income associated with years of school completion and work experience. One of the problems arising from the Mincer equation using the OLS (Ordinary Least Square) method is biased estimation result. The magnitude of the bias collected by previous researcher’s ranges from 0.7% - 9.4%. Bias can occur due to sample selection and endogeneity problems. Problem of sample selection will be overcome with mills ratio or Heckit method while the endogeneity problem in this paper will be overcome by IV (Instrumental Variable) method. Instrumental variables must have a provision: the variables used as the instrument have no relation with the dependent variable, and the variable has correlated with the independent variable that is considered endogenous. Variable used as instrument is parent education. However, in this paper, I use pooling data from 1993-2014 with Indonesia Family Live Survey. Empirical results show the difference between return on education value using OLS and IV methods. Results with IV give a larger return value than OLS. The return value bias for the Indonesian case is still in tolerance.

Keywords—return on education; mincer; pooling data; sample selection; instrumental variable

I. INTRODUCTION

Return to education is an economic advantage of one's investment in education. There are three different sides in defining the return of education, namely: private return, social return, and labor productivity return [1]. The famous return to education model is the Mincer model [2]. This paper focuses on private return and uses the Mincer equation to calculate the estimated return on education.

The Mincer equation allows for bias due to selection, unobserved ability, and endogeneity. So the education coefficient can’t be interpreted as a pure measure of educational impact. The coefficient of educational estimate becomes \( E(\hat{\beta}_e) \neq \beta_1 \) [3]. Some ways to overcome the bias include adding proxy variables to the equations, estimating models using twin data, using two-stage estimates, and using instrument variables [4].

A study using twins’ data provides estimation results that minimize bias, while when using IV (Instrumental Variable) method the estimation results obtained are greater than the estimation results using the OLS (Ordinary Least Square) method. In this study the instruments used were family backgrounds or the school system [5].

Until now there hasn’t been research that can conclude the magnitude of the bias that occurs due to using the OLS method in the Mincer equation [6]. Nevertheless, several studies have collected the results of previous studies on estimation of return on education using m IV method. The summary collected by the researchers provides an overview of the magnitude of the bias ranging from 0.7% - 9.4% [5,7].

In general, the estimation results carried out by several studies using IV method have an estimation value that is greater than the OLS method [8-13].

Studies on the return on education case of Indonesia have been carried out by several researchers. One researcher used a two-step regression method with a multinomial logit used in the first step, then the estimation results were included in the main model. This study uses IFLS1 to IFLS3 [14]. Other study used OLS and Heckit methods. It used SUSENAS 2005 data [15]. Other study used OLS, Heckit and fixed effect methods. Its study results that estimation of return on education with the Heckit and the Fixed Effect methods are smaller than the OLS method [16]. Other study used OLS as its basic methodology and compares it with two estimation steps from Heckman (Heckit's method) [17].

The benefits of this research are to find solutions to problem solving of the return on education estimation of the Mincer model for the case in Indonesia. In addition, to determine the amount of bias that occurs due to the use of OLS. Data for twins in Indonesia have not been collected well, so this paper uses the IV method to overcome the bias due to unobserved ability and endogeneity. The Heckit method will be used to see if there is a bias selection in this case.

However, the difference in the settlement of this case with previous researchers is the use of IV methods and the use of pooling cross-section data. It is hoped that this paper will provide additional information on the large bias that occurs and also shows the impact of time on changes in the value of return on education.
II. Method

A. Framework Mincer Equation

The simple model of school decision on partial balance describes the tradeoff in human capital investment. The framework of the income and education equation model refers to Acemoglu and Autor [18]. Every human being will try to maximize the utility function \( u(c) \) along the planning horizon \( T (T = \infty) \) with the positive discount rate \( (\rho > 0) \), and the rate of constant death rate \( (\nu \geq 0) \), which is reflected in the equation:

\[
\max_{0}^{T} \int \exp(-(\rho + \nu)t) u(c(t)) dt.
\]

Individuals have the human capital growth function \( h(t) = g_{h}(t)h(t), s(t) \), where each individual is within the interval \( S, s(t) = 1 \) at school, and \( s(t) = 0 \) when finished school. At the end of the school interval, the individual will have school level \( h(S) = \eta(s) \), where \( \eta(\cdot) \) is an increasing, continuous and concave function.

When \( t \in [s, \infty) \), the accumulation of human capital throughout life (due to individual work) has the equation \( h(t) = g_{h}(t)h(t) \), with the growth of human capital \( (g_{h} \geq 0) \). Individuals have an exponential wage growth with the equation \( w(t) = g_{w}(t)w(t) \), with growth wages \( (g_{w}) \) and wage conditions at the beginning \( w(0) > 0 \).

Suppose that \( g_{w} + g_{h} < r + \nu \), optimal school decision is a maximal solution of \( \max_{S} \int \exp(-(r + \nu)t) w(t)h(t)dt \). The equation is equivalent to

\[
\max_{S} \frac{\eta(S)w(0) \exp(-(r + \nu-g_{w})S)}{r + \nu - g_{h} - g_{w}}.
\]

First order condition equation (2) to \( S \) is

\[
\frac{\eta'(S^{*})}{\eta(S^{*})} = r + \nu - g_{w}.
\]

Equation (3) shows a high value at \( r \) (interest rate) and \( \nu \) (short time horizon) will decrease human capital investment. While the high growth value of wages increases the value of human capital and will encourage even greater investment.

By integrating both sides of equation (3) with respect to the variable \( S \) we obtain

\[
\ln \eta(S^{*}) = \text{constant} + (r + \nu-g_{w})S^{*}
\]

Earnings of a worker in the age of \( \tau \geq S^{*} \) in the labor market at time \( t \) will be obtained:

\[
W(S, t) = \exp(g_{w}t) \exp(g_{h}(t-S)) \eta(S).
\]

By providing the logarithm of equation (5) and substituting in equation (4) the income of the worker will be in the form:

\[
\ln W(S^{*}, t) = \text{constant} + (r + \nu-g_{w})S^{*} + g_{w}t + g_{h}(t-S^{*}),
\]

Where \( (t - S^{*}) \) is the worker's experience (time after school). If we use cross sectional to compare between workers, then time trend \( g_{w}t \) will be constant, so we get a canonical Mincer equation in which log wage is proportional to school and experience. Equation (6) can be simplified to be:

\[
\ln W_{i} = \text{constant} + \beta_{0}S_{i} + \beta_{1}E_{i}.
\]

Where \( i \) is an individual i. The coefficient value of \( \beta_{0} \) will be positive if \( r + \nu > g_{w} \). It will produce the coefficient \( \beta_{1} < \beta_{2} \). Experience variable \( (E) \) is individual potential experiences derived from calculations (age - years of education - 7).

B. Empirical Strategy

The Mincer model used is shaped

\[
\ln W_{i} = \beta_{0} + \beta_{1}S_{i} + \beta_{2}E_{i} + \beta_{3}E_{i}^{2} + \gamma X + \varepsilon.
\]

where \( X \) is a variable vector of individual characteristics.

This study uses pooling cross section. The advantages of using pooling cross-sections method are increasing sample size, and resulting in more precise estimates. Other benefits can identify the impact of time, so as to see whether the relationship between education and income changes over time [3]. The model pooling cross section is in the form of:

\[
\ln W_{i} = \beta_{0} + \beta_{1}S_{i} + \sum_{n=1}^{N} \beta_{3}S_{i} + \beta_{3}E_{i} + \beta_{3}E_{i}^{2} + \sum_{n=1}^{N} \gamma_{n}X_{i} + \varepsilon.
\]

The sample used in this research is the income workers. The selection of samples like this of course tends to the problem of sample selection. The problem of sample selection can be overcome by a two stage method introduced by Heckman [19]. The procedure for completion of sample selection correction as follows: (1) using all observations to find the probability of working by using probit model, then count the inverse Mills ratio for each observed sample. (2) Use the selected sample by entering the mills ratio value into the main model. Significant mills ratio coefficients indicate a selection problem bias, so the estimation of the coefficient of the main model with OLS will be biased [3]. This two stage method is known as the Heckit method.

Instrumental variable methods are used to overcome possible bias due to endogeneity. The problem of the instrument variable method is to find the right instrument for the education variable that is considered endogenous. Some studies used family backgrounds as instruments, such as: mother’s education [5], father's education or the education of
her siblings [20]. Other instruments used for educational variables include "birth quarter interacted with year of birth" [8,9], "closest distance to campus" [10], "age cohort indicator" [21], "dummy reform of the school system at the age of 13 years" [22]. In this paper, I use parent education (father or mother) as an instrument of the educational variables.

III. RESULTS

This paper uses IFLS data from first wave (1993) to fifth wave (2014). The respondents selected who have ages between 15 and 65 years. The number of observations in this data is shown in Table 1. All observations after deduction of education missing and ages limits totaled 114,618 and those with income data were 68,060. Decrease in observation due to missing data on one's income.

The statistical summary of the data obtained is shown in Table 2. The average education in Indonesia increases from year to year. The average education in 1993 was 5,365 years and in 2014 it was 9,043 years.

The probability model working or not working from all respondents needs to be made first. The descriptive analysis of the probability model is shown in table 3. This probability data is used to estimate the Probit model to overcome the selection bias.

The average probability of work increases every year except in 1997. The average probability of employment in 2014 increased by around 17.2% compared to 1993. While in 1997 it decreased by around 9.6% compared to 1993.

### Table I. Summary of Observations

| Observation Dumped | Remainder |
|--------------------|-----------|
| IFLS1 | | |
| Individuals who answered Book III | 14,418 |
| Observation discarded due | |
| Education missing | 21 | 14,397 |
| Age > 65 | 1,248 | 13,149 |
| Income missing and outlier | 5,717 | 7,432 |
| IFLS2 | | |
| Individuals who answered Book III | 19,910 |
| Observation discarded due | |
| Education missing | 45 | 19,865 |
| Age > 65 | 1,264 | 18,601 |
| Income missing and outlier | 9,096 | 9,505 |
| IFLS3 | | |
| Individuals who answered Book III | 25,490 |
| Observation discarded due | |
| Education missing | 26 | 25,464 |
| Age > 65 | 1,666 | 23,798 |
| Income missing and outlier | 9,954 | 13,444 |
| IFLS4 | | |
| Individuals who answered Book III | 29,059 |
| Observation discarded due | |
| Education missing | 3 | 29,056 |
| Age > 65 | 1,734 | 27,322 |
| Income missing and outlier | 11,186 | 16,135 |
| IFLS5 | | |
| Individuals who answered Book III | 34,464 |
| Observation discarded due | |
| Education missing | 33 | 34,431 |
| Age > 65 | 2,683 | 31,748 |
| Income missing and outlier | 10,604 | 21,144 |

Source: IFLS1, IFLS2, IFLS3, IFLS4, and IFLS5

### Table II. Summary of Work Probability Model

| Data | Variable | Observation | Average |
|------|----------|-------------|---------|
| IFLS1 | Worked Probability | 13,149 | 0.565 |
| | Education | 3,365 | 0.28 |
| | Age | 40.074 | 0.7%
| | Men | 0.433 | 0.347 |
| IFLS2 | Worked Probability | 18,601 | 0.511 |
| | Education | 6,755 | 0.457 |
| | Age | 34.870 | 0.475 |
| | Men | 0.436 | 0.355 |
| IFLS3 | Worked Probability | 23,798 | 0.582 |
| | Education | 7,484 | 0.36 |

Source: IFLS1, IFLS2, IFLS3, IFLS4, and IFLS5

Table 2. Cont.

| Data | Variable | Observation | Average |
|------|----------|-------------|---------|
| IFLS4 | Worked Probability | 27,321 | 0.569 |
| | Education | 8,367 | 0.429 |
| | Age | 34.594 | 0.617 |
| | Men | 0.477 | 0.398 |
| IFLS5 | Worked Probability | 31,748 | 0.662 |
| | Education | 9,043 | 0.515 |
| | Age | 35.617 | 0.617 |
| | Men | 0.480 | 0.403 |

Source: IFLS1, IFLS2, IFLS3, IFLS4, and IFLS5
Based on the highest level of education achieved, someone who is educated above high school has the highest probability of working, followed by elementary education level. The lowest probability is achieved when a person’s highest education is junior high. This is possible because at the age of 15-18 years some individuals still take high school education. Elementary education level in addition to having a high probability of work also has the largest frequency of around 37% of all data. This indicates that education is still quite low, especially if the added respondents who did not finish elementary school to about 45.9%. This condition also shows that workers with elementary education are still quite high in Indonesia.

Based on gender, men who worked were around 77.66% while women were around 42.88% (Table 3). The number of women who work has experienced a significant increase compared to men. The average education and average income of workers has increased from year to year. Women's education has increased higher than men (Table 4), as well as for the average income (Table 5). This shows a reduction in the inequality of education and income between men and women.

### TABLE III. DESCRIPTIVE ANALYSIS OF WORK PROBABILITY 1993-2014

| Independent Variable | Probability | No Worked | Worked | Total |
|----------------------|-------------|----------|--------|-------|
|                      | freq | %     | freq | %     | Freq | % |
| Year                 |      |       |      |       |      |   |
| 1993                 | 5,717 | 43.48 | 7,432 | 56.52 | 13,149 | 100 |
| 1997                 | 9,096 | 48.90 | 9,505 | 51.10 | 18,601 | 100 |
| 2000                 | 9,954 | 41.83 | 13,844 | 58.17 | 23,798 | 100 |
| 2007                 | 11,187 | 40.95 | 16,135 | 59.05 | 27,322 | 100 |
| 2014                 | 10,735 | 33.81 | 21,013 | 66.19 | 31,748 | 100 |
| Education            |      |       |      |       |      |   |
| No school            | 4,627 | 45.78 | 5,481 | 54.22 | 10,108 | 100 |
| PS                   | 16,003 | 37.68 | 26,467 | 62.32 | 42,470 | 100 |
| JHS                  | 10,667 | 48.78 | 11,201 | 51.22 | 21,868 | 100 |
| SHS                  | 12,380 | 41.61 | 17,373 | 58.39 | 29,753 | 100 |
| College              | 3,012 | 28.91 | 7,407 | 71.09 | 10,419 | 100 |
| Cohort               |      |       |      |       |      |   |
| Young                | 24,440 | 55.51 | 19,586 | 44.49 | 44,026 | 100 |
| 15-29                | 8,795 | 29.89 | 20,628 | 70.11 | 29,423 | 100 |
| 30-39                | 3,464 | 26.79 | 14,933 | 73.21 | 20,397 | 100 |
| Old                  | 7,990 | 38.47 | 12,782 | 61.53 | 20,772 | 100 |
| Gender               |      |       |      |       |      |   |
| Men                  | 12,060 | 22.34 | 41,932 | 77.66 | 53,992 | 100 |
| Female               | 34,629 | 57.12 | 25,997 | 42.88 | 60,626 | 100 |
| Status               |      |       |      |       |      |   |
| Married              | 28,459 | 34.99 | 52,874 | 65.01 | 81,333 | 100 |
| No                   | 18,230 | 54.77 | 15,055 | 45.23 | 33,285 | 100 |

Based on the results of various estimates using the OLS, Heckit and IV methods are shown in Table 6. The return value of 15.2% is the return value in 1993 for the OLS method. The estimated return in 1997 declined 1.8% compared to 1993 to 13.4%. In 2000 decreased 3.2% compared to 1993 to 12%. In 2007 decreased by 2.9% compared to 1993 to 12.3%. In 2014 decreased 4.5% compared to 1993 to 10.7%.

### TABLE IV. AVERAGE YEAR OF WORKERS EDUCATION IN INDONESIA

| Year | Average Year of Education |
|------|---------------------------|
|      | Men | Female | Total |
| 1993 | 6.16 | 4.88 | 5.72 |
| 1997 | 6.99 | 6.19 | 6.68 |
| 2000 | 7.76 | 6.78 | 7.38 |
| 2007 | 8.62 | 8.13 | 8.44 |
| 2014 | 9.28 | 9.06 | 9.19 |

### TABLE V. AVERAGE REAL INCOME OF WORKERS IN INDONESIA

| Year | Average Income (Rp) |
|------|---------------------|
|      | Men | Female | Total |
| 1993 | 4,525,068 | 2,781,745 | 3,929,261 |
| 1997 | 4,777,493 | 3,074,796 | 4,128,121 |
| 2000 | 4,554,587 | 2,947,824 | 3,939,690 |
| 2007 | 5,737,948 | 4,010,764 | 5,101,132 |
| 2014 | 7,799,850 | 6,941,581 | 7,448,773 |

Source: IFLS1, IFLS2, IFLS3, IFLS4, and IFLS5

The estimate of return value with Heckit method is 12.7% in 1993. There was a decrease of 1.79% in 1997 compared to 1993 to 10.91%. In 2000 there was a decrease in return of 3.26% compared to 1993 to 9.44%. In 2007 there was a decrease of 2.94% compared to 1993 to 9.76%. A decline of 4.55% occurred in 2014 compared to 1993 to be 8.15%.

The author used family background instruments that are father education and maternal education. Both instruments result in a statistical test of F greater than 10 for the first stage regression, so the instrument is feasible for use in method IV. The estimated value of educational coefficients and work experience for method IV has a greater value than the OLS method.
IV. DISCUSSION

The estimation results of the return on education value obtained by the author are quite large especially when compared to developed countries. But these results support the findings of Psacharopoulos. These are relative differences in human capital in developing countries compared to developed countries [23]. The level of return on education services for developing countries is higher than that of developed countries [23,24]. This result is also consistent with a summary of several studies on the value of educational services from Bils and Klenow, which most of the estimated returns for developing countries are higher than developed countries [25].

The Heckit method is used to overcome bias selection in OLS models. It appears that Heckit's estimation method is lower than OLS, which denotes the bias due to sample selection. The bias size of this sample selection is around 2.5%. These results differ from the findings of Purnastuti et al. who did not find enough evidence of bias due to sample selection. The educational variables used dummy sets of graduates of educational level [17]. This is in contrast to my study who use the year of education as a measure of educational variables.

The IV method in this paper produces a higher estimate of OLS. The difference in the estimated return on education value is around 6%. This larger value of IV estimates is consistent with most of the previous researchers [5]. The difference in return on education is still within the range of empirical study results the previous studies [5,7]. Card concluded that estimates with IV provide higher values above 30%. The estimate IV method is greater due to the trend of estimating the average effect in the low group. This term is known as LATE (Local Average Treatment Effect) [26].

ACKNOWLEDGMENT

Thank you to the Universitas Sangga Buana institution for giving support in this research. I do not forget to thank the colleagues who supported me to write this article.

REFERENCES

[1] R. Blundell, L. Dearden, and B. Sianesi, “Estimating the returns to education: Models, methods and results”, 2001.
[2] J. Mincer, Schooling, Experience, and Earnings. New York: Columbia University Press, 1974.
[3] J. Wooldrige, “Introductory econometrics: A modern approach”, Cengage Learning, 2012.
[4] C. Harmon, and I. Walker, “Estimates of the economic return to schooling for the United.” The American Economic Review, vol. 85, no. 5, pp. 1278-1286, 1995.
[5] D. Card, “The causal effect of education on earnings,” Handbook of labor economics, vol, 3, pp. 1801-1863, 1999.
[6] D. Card, "Estimating the Returns to Schooling: Progress on Some Econometric Problems," Econometrica, vol. 69, pp. 1177-1180, 2001.
[7] C. Harmon, and H. Osterbeek, "The Returns to Education: Microeconomics," Journal of economic surveys, vol. 17, no. 2, pp. 115- 156, 2003.
[8] J.D. Angrist and A.B. Krueger, “Does Compulsory School Attendance Affect Schooling and Earnings?” The Quarterly Journal of Economics, vol. 106, no. 4, pp. 979-1014, 1991.
[9] D. Stauger, and J.H. Stock, “Instrumental variables regression with weak instruments”, Econometrica, vol. 65, pp. 557-586, 1997.
[10] T.J. Kane and C.E. Rouse, "Labor market returns to two- and four-year colleges: is a credit and do degrees matter?" Working paper no. 4268 (NBER, Cambridge, MA), 1993.
[11] D. Card, “Earnings, ability and schooling revisited,” Research in Labour Economics, vol. 14, 1995.
[12] E. Duflo, “Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment,” The American Economic Review, vol. 91, no. 4, 2001.
[13] K.J. Denny, and C.P. Harmon, “Testing for sheepskin effects in earnings equations: evidence for five countries,” Applied Economics Letters, vol. 8, no. 9, pp. 635-637, 2001.
[14] G.D.V. Pattinasarany, “Estimating returns to schooling in Indonesia: Evidence from the Indonesia Family Life Survey, 1993–2000.” 3116017, Michigan State University, Ann Arbor.” 2003, [Online] Retrieved from http://search.proquest.com/docview/305328333?accountid=13771
[15] Anggraini, “Pengaruh Gender pada Tingkat Pengembalian Investasi Pendidikan,” 2007.
[16] Mustofa, “Return to Education Tenaga Kerja di Indonesia: Analisis Data IFIS 2000 dan 2007,” Unpublished, 2011.
[17] L. Purnastuti, P.W. Miller and R. Salim, “Declining Rates of Return to Education: Evidence for Indonesia,” Bulletin of Indonesian Economic Studies, vol. 49, no. 2, pp. 213-236, 2013.
[18] D. Acemoglu, and D. Autor, “Lectures in Labor Economics,” Unpublished.
[19] J.J. Heckman, “Sample Selection Bias as a Specification Error,” Econometrica, vol. 47, no. 1, pp. 155-162, 1979.
[20] O. Ashenfelter, and D.J. Zimmerman, “Estimates of the returns to schooling from sibling data: Fathers, sons, and brothers,” Review of Economics and Statistics, vol. 79, no. 1, pp. 1-9, 1997.
[21] A. Ichino and R. Winter-Ebner, 'The long-run educational cost of World War II', Unpublished discussion paper (European University Institute), 1998.
[22] C. Meghir and M. Palme, “Assessing the Effect of Schooling on Earnings using a social experiment,” IFS working paper no. W99/10, 1999.
[23] G. Psacharopoulos, ‘Returns to Investment in Education: A global Update,’ World development, vol. 22, no. 9, pp. 1325-1343, 1994.
[24] G. Psacharopoulos, ‘Returns to Education: an Updated International Comparison,’ Comparative education, vol. 17, no. 3, pp. 321-341, 1981.
[25] M. Bils and P.J. Klenow, “Does Schooling Cause Growth?” American Economic Review, pp. 1160-1183, 2000.
[26] P. Oreopoulos, “Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter,” The American Economic Review, vol. 96, no. 1, pp. 152-175, 2006.