IBS
POPULATION PROGRAM ■
INSTITUTE OF BEHAVIORAL SCIENCE ■
UNIVERSITY OF COLORADO AT BOULDER ■

WORKING PAPER

The Return to Education in Thailand: A Pseudo-Panel Approach

Sasiwimon Warunsiri
Robert McNown

June 2009

Population Program POP2009-02

The Return to Education in Thailand: A Pseudo-Panel Approach

Sasiwimon Warunsiri
Department of Economics
University of Colorado at Boulder
School of Economics
University of the Thai Chamber of Commerce, Thailand

Robert McNown
Department of Economics
Institute of Behavioral Science
and
Program on International Affairs
University of Colorado at Boulder

June 2009

Correspondence address: Robert McNown, Department of Economics, University of Colorado at Boulder, 256 UCB, Boulder, Colorado, 80309-0256, USA. Tel.: +1-303-492-8295; Fax: +1-303-492-8960; Email-address: mcnownr@colorado.edu

Acknowledgements: The data used in this paper is provided by The University of Chicago-UTCC Research Center at The University of the Thai Chamber of Commerce, Thailand. This research has benefited from the NICHD-funded University of Colorado Population Center (grant R21 HD51146) through administrative and computing support. The content is solely the responsibility of the author and does not necessarily represent the official views of NIH or NICHD.

The Return to Education in Thailand: A Pseudo-Panel Approach

Abstract:

This study employs the pseudo-panel approach for estimating the rate of return to education in Thailand, avoiding the omitted variable bias common to estimates from data on individuals. Pseudo-panel data are constructed from repeated cross sections of Thailand's National Labor Force Surveys (1986-2005) of workers who were born between 1946 and 1967. One fundamental finding is that there is a downward bias of the rate of return to education in Thailand in cross-sectional least squares regressions on the individual data, contrary to the usual positive bias found in U.S. data. This bias is due to the omitted ability variable from the Mincerian equation that is negatively correlated with years of education, due to a high opportunity cost of education for high ability workers. The overall rate of return is estimated to be between 14% and 16%. Females have higher returns than males, and the return for unmarried workers exceeds that of married workers.

JEL Classification: C23; I21; J24
Keywords: Returns to education; Pseudo-Panel; Synthetic cohort; Thailand; Asia

1 Introduction

The rate of return to education has been widely studied since the late 1950s. The conventional approach used to estimate the return to education is the standard Mincerian earnings function, introduced by Jacob Mincer (1974). Even though hundreds of papers studied this issue in different countries, different time periods, and different estimation methods[1], few studies produce the "true" rate of return to education (Heckman, Lochner, and Petra, 2005, p.3).

The main problem is that an unobserved "ability" variable, which correlates with years of education and with earnings, is omitted from the estimated Mincerian earnings function. As a result, the coefficient on years of education from least squares regressions on individual data suffers from "ability bias" (Card, 1999). Given the expected positive correlations between ability and both earnings and years of schooling, the standard critique emphasizes an upward bias. However, the correlation between ability and education could be negative, or there may be other factors besides ability that could cause a bias of a different nature, possibly downward, as in other empirical studies (Ashenfelter et al, 1999).

The two conventional methods for correcting the bias are Instrumental Variables (IV) estimation and panel estimation with individual fixed effects. However, IV estimation faces the problem of selecting a valid IV. In particular, if the instrument positively correlates with earnings, the estimates can become even more upward biased (Ashenfelter et al, 1999). Panel

[1] According to Psacharopoulos and Patrinos (2004), there were ninety-eight countries, including both developed countries and developing countries, which have already estimated rates of return to education. The average rate of return to education across these studies is 10%. Also, this rate tends to be higher in the developing countries than in developed countries. Moreover, women in general tend to have a higher rate of return than men.

estimation with fixed effects can eliminate the bias caused by the different abilities across individuals, but the main limitation is the availability of the data, especially in developing countries that usually have only cross-sectional data.

This study employs a pseudo-panel approach as an alternative means for estimating the rate of return to education in Thailand, which is representative of small developing countries facing this data limitation. The pseudo-panel approach controls for unobserved ability or other individual specific effects that may otherwise bias the estimated rate of return to education in individual cross sectional regressions. By constructing a pseudo-panel (or synthetic cohort data set - Deaton, 1985) from repeated cross sectional surveys of Thailand's National Labor Force Survey (1986-2005), this paper presents estimates the rate of return to education for Thai workers who were born between 1946 and 1967.

Comparing the results from the pseudo-panel approach with cross-sectional regression estimates with individual data, the years of education coefficients indicate that there is a downward bias in cross-sectional estimates of the return to education in Thailand. This result holds when controlling for differences across cohorts and over years of the survey, and also for estimates with several disaggregations of the data by demographic characteristics.

The downward bias conflicts with the standard result, which implies positive associations between the omitted ability variable and both earnings and the level of education. Two alternative explanations for a downward bias in Thai data are considered, namely, aggregation bias and an opportunity cost argument. However, when the data are disaggregated according to alternative demographic characteristics, the downward bias remains in the individual data

estimates. Alternatively, the opportunity cost argument states that high ability individuals have high wage options by entering the labor force, implying high opportunity costs of schooling. This could lead to a negative correlation between ability and years of education in the Thai data. Combined with a positive correlation between ability and earnings, the downward bias in the individual data estimates emerges.

This paper presents unbiased estimates of the returns to education for Thai workers using the pseudo-panel approach. The following section discusses related literature. Section 3 describes the pseudo-panel methodology used in the estimation. Section 4 presents the synthetic cohort data set and the variables used in the estimation. The results and discussion are in section 5, and the paper concludes in section 6.

2 Related Literature

The foundation for estimating the rate of return to education was developed by the specification of Jacob Mincer (1974). Setting the logarithm of earnings as the dependent variable, the number of years of schooling as an independent variable, and controlling for the number of years of experience and other individual characteristics, the years of schooling coefficient from ordinary least squares (OLS) regression is interpreted as the rate of return to education.

Even though the Mincerian model is a standard method for estimating the rate of return to education, it suffers from the problem of omitted variables, particularly the unobserved ability

variable, causing OLS estimators to be biased. Griliches (1977, p.4) states that the schooling coefficient from the least squares estimator is biased upward based on these three main assumptions: (1) the ability variable positively correlates with earnings, (2) the excluded ability variable positively correlates with the schooling variable, and (3) the ability variable is the only variable that is excluded. Alternatively, if unobserved ability is negatively correlated with years of schooling and positively correlated with earnings, we can expect downward biased least squares estimates (Kalwij, 2000).

Some studies take ability into account in the estimation by employing various instrumental variables such as the quarter of birth (Angrist and Krueger, 1991) and distance to school (Kane and Rouse, 1993). However, Bound, Jaeger and Beker (1995) found that the results from IV estimation become less accurate than the OLS estimation. Card (1999) and Card and Lemieux (2001) conclude that IV estimates of the rate of return to education will be higher or lower than OLS estimates depending on what types of instrumental variables are used.

Panel data estimators are also employed, to control for unobserved individual effects, for example, in the study by Harmon and Walker (1995) that uses data on men from the British Family Expenditure Survey. However, panel wage data are often not available, particularly for developing nations. Consequently, individual cross sectional data is most commonly used in the estimation of returns to education in developing countries, even though there is reason to question whether the estimate can reflect the “true” rate of return to education.

Regarding studies of the return to education in Thailand, Chiswick (1976) first introduces an estimation of the earnings function in Thailand as a case study of developing countries. In

addition to a regression on the Mincerian model, the paper develops the technique for analysis of earnings by the self-employed workforce. This relates to imputed income of self-employed workers as wages. One finding is that the estimated coefficient on schooling for women is higher than for men.

Amornthum and Chalamwong (2001) update the rate of return to education in Thailand in 2000 using the framework of the World Bank, applying OLS to the basic Mincerian equation, but adding dummy variables such as location and marital status as controls. Contrary to Chiswick they find that the rate of return to education is higher for men than for women.

The most recent study is conducted by Hawley (2004) who studies the effect of the macroeconomy on the return to education in three different years (1985, 1995, and 1998). The results show the rate of return to education is stable across time and across gender.

The main theme of these studies is to find the rate of return to education in Thailand in different time periods using the cross-sectional analysis. However, aside from the problem of unobserved individual heterogeneity, Glenn (2005, p.3) points out another weakness of using cross-sectional data: "The difference by age shown by cross-sectional data may or may not be age effects, because people of different ages are members of different cohorts and may have been shaped by different formative experiences and influences."

In other words, individual workers in different cohorts have different opportunities, attitudes, and behaviors. For example, the availability of schooling, as well as the quality of the school as a result of technological changes that lead to a different quality of training, varies over time. In addition, average wages observed at different survey years may vary with

macroeconomic conditions. As a result, different cohorts will earn different average wages at various points in time, so that estimates of rates of return to education will vary across cohorts and over time. This points to the necessity of controlling for cohort specific and time specific effects in the pseudo-panel analysis.

The previous studies for Thailand fail to solve the problem of omitted variables bias (the issue of the unobserved heterogeneity). Nor do they control for differences across cohorts or time that may also bias the estimates of the rate of return to education. Therefore, a re-examination of the return to education in Thailand, as representative of small and open developing economies, is in order. Towards this end this study builds synthetic cohorts to deal with problems of unobserved heterogeneity, controlling also for cohort and time specific effects to produce unbiased estimates of the rate of return to education in Thailand.

3 Methodology

This study begins with the basic "Human Capital Earnings Functions" (Mincer, 1974):

$$\ln w = \alpha + \beta_0 E + \beta_1 X + \beta_2 X^2 + \varepsilon \qquad (1)$$

where *ln w* is the natural log of the hourly wage, *E* is the number of years of education, and *X* is the number of years of experience (or age). Equation (2) is the time, year, and individual specific representation of equation (1), where *i* indexes individuals (*i =1,…,N*), *c* indexes cohorts (*c =1,…,C*), and *t* indexes time periods *(t = 1,…,T)*.

$$\ln w_{ict} = \gamma + \beta_1 E_{ict} + \beta_2 X_{ict} + \beta_3 X_{ict}^2 + \delta_i + \eta_{ct} + u_{ict} \qquad (2)$$

Here δ_i captures individual effects (such as different abilities across individuals) and η_{ct} captures cohort-year effects (due to differences in macroeconomic conditions or education quality at different years and for different cohorts). Let $\alpha_{ict} = \delta_i + \eta_{ct}$. Although it is assumed that u_{ict} is uncorrelated with E_{ict}, X_{ict}, and α_{ict}, δ_i and η_{ct} are likely to be correlated with E_{ict}. It is not possible to include the "ability" variable into the equation or directly use individual fixed effects for controlling unobserved individual heterogeneity when estimating (2) with individual survey data, so that least squares estimates of (2) will be biased and inconsistent.

To solve this problem, Deaton (1985) defines a set of *C (c=1,…,C)* cohorts, based on year-of-birth. Averaging (2) over the cohort members eliminates the individual heterogeneity (δ_i), such as the differing abilities across individuals, leaving $\eta_{ct} = \overline{\alpha}_{ct}$.

$$\overline{\ln w_{ct}} = \beta_1 \overline{E_{ct}} + \beta_2 \overline{X_{ct}} + \beta_3 \overline{X_{ct}^2} + \overline{\alpha_{ct}} + \overline{u_{ct}} \tag{3}$$

In addition, inclusion of cohort dummies (f_c) and year dummies (f_t) extracts time and cohort effects from the error term, leaving only the idiosyncratic error, $\overline{u_{ct}}$.

$$\overline{\ln w_{ct}} = \beta_1 \overline{E_{ct}} + \beta_2 \overline{X_{ct}} + \beta_3 \overline{X_{ct}^2} + f_t + f_c + \overline{u_{ct}} \tag{4}$$

Estimation of (4) is based on cohort means for each year. For example, $\overline{\ln w_{ct}}$ is the average of $\ln w$ over the sample observations in cohort *c* at time *t*. In (4) all error components in (2) that are correlated with explanatory variables have been purged from the error term, so that fixed effects estimation of this equation expressed in terms of cohort means is consistent. Not only does

estimation of (4) deal with problems of individual heterogeneity while controlling for year and cohort effects, the use of cohort means can "average out" individual measurement errors (Antman and McKenzie 2007)

The remaining concern of this approach is the possibility of biases in cases of small group sizes (Deaton 1985). The reason is that the cohort means ($\overline{\alpha}_{ct}$) are not constant temporally and may differ from the true cohort mean because the observations within each cohort are collected at different points in time; thus, $\text{cov}(\overline{\alpha_{ct}} - \alpha_c, \overline{E_{ct}}) \neq 0$ in small samples, where α_c is the true cohort effect (Devereux 2007).

Some studies find that the "sampling error" problem will not occur with at least 100-200 observations per cell (Verbeek and Nijman, 1992, 1993). However, Devereux (2006) contends that a larger number of observations per cohort-year group may be necessary to avoid substantial biases, and this can be achieved by grouping cohorts. In this study estimates are presented based on a pseudo-panel data set with one-year cohorts and another with two-year cohorts to check the sensitivity of estimates to cell sizes. In addition, the pseudo-panel estimates are compared with estimates from a regression on individual data to see the effects of controlling for individual heterogeneity. Finally, to control for biases arising from inappropriate aggregation of the data, the estimates from the full sample are supplemented with results from samples disaggregated by demographic characteristics including gender, rural/urban residence, and marital status.

Aggregation bias can be viewed as a form of omitted variables bias if returns to education differ across demographic groups. If marital status, for example, affects wages and is also correlated with years of schooling, then failure to control for marital status imparts omitted

variables bias. Disaggregating the pseudo panel data into married and non-married groups allows flexibility in the estimated effect of marital status on the relation between wages and education.

4. Data and variables

Construction of a pseudo-panel (Deaton 1998) starts by using the age of each individual at the time of the survey to establish the birth cohort to which they belong. For every survey year the individual observations on the variables of interest are then averaged across each birth cohort, creating cohort-year averages as the units of observation. Cohorts are defined for birth years from 1946 to 1967 using data from surveys for 1986 through 2005. This establishes age 19 (e.g., in 1986 from the first birth cohort) as the youngest individuals in the sample.

There are 199,833 individual observations from which to build the pseudo-panels. The first data set pools data from 22 single year-of-birth cohorts and 20 survey years for a total of 440 cohort-year observations. In every case cell sizes exceed 100, and the vast majority contain over 200 individuals (Appendix A). The two-year cohorts are summarized in Appendix B. Only two cells contain fewer than 300 individuals. In this pseudo-panel the total number of observations available for the estimation is 220 cohort-year groups (= 11 cohorts x 20 years of survey). Additional pseudo-panels are defined from disaggregations according to gender, place of residence, and marital status using the two-year cohort design in order to maintain adequate cell sizes.

The data were collected by the National Statistical Office of Thailand (NSO), Statistical Forecasting Bureau, as part of the National Labor Force Surveys (LFS) for 1986-2005. Each quarterly LFS represents data compiled from interviews with the head of household or members

of household, with 70,000-200,000 people representing 0.1-0.5% of the total Thai population. For the year 1985-1999, data are available for only the first and third quarters, but from 2000, the NSO began collecting data every quarter. This study employs third quarter data in the estimation in order to control for the effect of seasonal agricultural labor movement. Thai agricultural workers migrate to work in the cities during the dry season, but return home during the rainy season of the third quarter (Chalongphob and Yongyuth, 1996). The sample is limited to people whose working hours are equal to or greater than 30 hours a week, and those of ages 19-59 at the time of each survey. This sample design eliminates individuals who might be working part-time while still in school or partially retired.

The three primary variables of this study are hourly wages, years of education, and age. The hourly wage is constructed from the monthly wage recorded in the survey using the reported number of hours of work.[2] This nominal wage is deflated by the Thailand Consumer Price Index (CPI)[3]. The LFS records the highest attained degree, and these data are converted into years of education ranging from zero (no education) to 23 years for those with PhDs. Age is reported directly in the LFS, and this variable is entered into the regressions in both linear and squared terms.

---

[2] Welsh (1997) discuss the problem of constructing hourly wages from annual earnings, weeks, and hours per week in the estimation of the responsiveness of labor supply to hourly wage rates. A problem of "division bias" can arise with errors in reporting hours of work when both dependent and independent variables involve this noisy measure (Borjas, 1980). In this study, however, wages only appear as the dependent variable, avoiding this concern. The hourly wage is constructed from the monthly wage dividing by 4 to obtain weekly wage and further dividing by reported weekly hours to obtain the hourly wage.

[3] The CPI indexes (2002 as a base year) are from the Bureau of Trade and Economic Indices, Ministry of Commerce, Thailand

5 Results and discussion

5.1 Aggregated Estimates

The estimates from the regressions with individual data, one-year cohort means, and two-year cohort means are presented in table 1. Column (1) shows the results from cross-sectional regression on individual data, and the estimates from the pseudo-panel method are presented in columns (2)-(5). Year dummies are included in all five cases reported in Table 1, and cohort dummies are added for the results in columns (3) and (5). Columns (2)-(3) show the result from one year cohort means, to compare with the estimates based on two-year cohorts. Although the cell sizes in the latter case exceed 283 vs. only 112 for the single year cohorts, the similarities between these two sets of estimates indicate no apparent biases with the smaller cell sizes. This evidence is consistent with Verbeek and Nijman (1992, 1993), who contend that 100 observations per cell is sufficient to avoid biases in a pseudo-panel estimation. Furthermore, comparisons between columns (2) and (3) and across (4) and (5) show that controlling for cohorts does not have an important effect on the estimates[4].

The basic finding of Table 1 is that the estimated returns to education from the pseudo-panels are considerably larger than those from regressions with individual data. When using the pseudo-panel approach, the years of education coefficient ranges between 0.145 and 0.161, compared with 0.115 from the individual regression. This latter estimate is in the 8% - 12% range of estimates in previous cross-sectional studies of Thai workers cited in section 2. The

[4] To check the robustness of the pseudo-panel design, the first and last cohorts are dropped from the sample, and the remaining cohorts are recombined into different two-year groupings. This results also in a change in sample size with 400 cohort-year observations constructed from 184,093 individual data points. With this new pseudo-panel the coefficients on years of education and other coefficient estimates are similar to those from the full sample.

robustness of the estimate across the four pseudo-panel estimations reinforces the conclusion of a downward bias in the individual data regressions, which contrasts with the upward bias that has been generally reported for US data.

In most returns to education studies using individual data, the usual expectation is that the coefficient on education will be biased upwards due to the omission of individual ability that is positively correlated with both education and earnings. However, an opportunity cost argument can give rise to a negative association between ability and education and thus account for the downward bias found here. Individuals with greater ability have high potential wages, representing a high opportunity cost of studying. As a result, high ability individuals may choose to work instead of studying, creating a negative correlation between ability and years of education. This effect may be strengthened if the direct costs of schooling are substantial and there is little opportunity to finance education by borrowing or intergenerational transfers.

5.2 Disaggregation by Gender

Given the unusual finding in the Thai data, other explanations for the bias in the individual data estimates are considered. In particular, to examine the possibility of aggregation bias, the sample is disaggregated across three alternative demographic dimensions: gender, place of residence, and marital status. Since disaggregation reduces the numbers of observations in each cell in the pseudo-panels, these disaggregated panels are constructed using two-year cohorts.

Table 2 shows the regression results of equation (4) with the disaggregated data set, which has been stratified by men and women. Overall, the results in table 2 confirm the main results in table 1, showing the downward bias in cross-sectional regressions on individual data.

The coefficients on years of education for men and women from the cross-sectional regression are 0.107 and 0.129, respectively (columns (1) and (4)), while, from the two-year cohort means they are around 0.12 for men and 0.16 for women. This disaggregation shows the rate of return to education for women is higher than for men. This result is consistent with the many studies of US, in which the rate of return to schooling for women is always greater than for men (Dougherty, 2005)[5], but contrasts with some studies for Thailand (see section 2). Dougherty's explanation of the higher rate of return for women is that education helps women find employment outside "the low-paying traditionally female occupations".

The downward bias in the returns to education estimate from the individual data regressions remains with disaggregation by gender. In addition, the difference between the individual data estimate and the pseudo-panel value is greater for women (0.04=0.16-0.12) than for men (0.02=0.12-0.10). Applying the opportunity cost argument presented above, this difference could mean that high ability men have greater educational access than women, due to the attitudes and conventions of Thai society during the 1950s-1960s. During this early period of development, there was discrimination against girls in education (Thosanguan, 1978). Gandhi-Kingdon (2002) defined this as "unexplained parental discrimination", with differential support for educating boys over girls. A girl with abilities equal to a boy's would receive less family support for schooling, thus strengthening the negative correlation between years of education and ability for girls compared with boys, and increasing the downward bias observed for females.

[5] Dougherty (2005) draws this conclusion from 28 US studies on the rate of return to education between men and women.

5.3 Rural VS Urban Disaggregation

Table 3 displays the individual and pseudo-panel estimates separately for urban and rural residents. Overall, these estimates are consistent with the main results in table 1, again showing the downward bias in cross-sectional regressions on individual data.

Focusing on the pseudo-panel estimates with cohort fixed effects, the coefficient on years of schooling is higher for those living in urban areas (0.158) compared with that for rural residents (0.136). This is consistent with the expectation that individuals living in urban areas have more opportunities to exploit skills acquired by higher education than do those living in rural areas.

The gap between pseudo-panel and individual data estimates of the returns to education indicate a slightly larger bias for urban versus rural workers. When cohort dummies are included, for example, the difference between estimates is four percent for urban residents and two percent for those in rural areas. Given higher relative wages in urban areas, the opportunity cost of studying for urban residents is higher than for those in rural areas. In addition, people living in rural areas may be able to work on their farms at the same time as studying, so that the opportunity cost of studying for rural areas is lower than for urban areas. These differences in opportunity costs may account for the greater downward bias in the estimate of the returns to education for urban compared with rural workers.

5.3 Diaggregation by Marital Status

The results of the individual data and two-year cohort mean regressions for the married group and the non-married group are presented in table 4. The regressions for married workers provide the first estimates that show no important differences between the pseudo-panel and the individual data regressions. In addition, there is the somewhat surprising result that the returns to education are higher for non-married workers than for married workers (16 percent versus 11 percent for the pseudo-panel regressions with cohort dummies).

It is difficult to understand how marital status might be related to levels of education to account for the differences in results between married and unmarried workers. One possible explanation is that some of the cell sizes are too small to achieve unbiasedness in the pseudo-panel estimates. However, only five cells in the married panel have fewer than 100 observations, which is the sample size found by Verbeek and Nijman (1992, 1993) to be adequate for avoiding biased estimates.

Alternatively, there could be unobserved characteristics of married workers that correlate with years of schooling or with wages so as to lower their estimated returns to education. Since the decision to leave school is generally taken prior to marriage, marital status is not a causal determinant of years of education. On the other hand, those leaving school earliest may be most likely to marry early, as they have begun to establish a record of work and income to support a family. Others who stay in school longer and remain unmarried may be geographically mobile and able to take on positions with higher risk-return profiles. These differences could account for the relatively lower returns to education for married workers that is shown in the pseudo-panel estimates.

The general conclusion from the disaggregated estimates is that the downward bias in the individual data estimates is not an aggregation problem. Rather, it can be explained by an opportunity cost argument. Individuals with greater ability have high potential wages, and therefore choose work instead of additional education. This could cause a negative correlation between ability and years of schooling, an effect that may be stronger if the direct costs of schooling are substantial and there is little support for education from the government or other sources.

6 Conclusion

This study applies a pseudo-panel approach to estimate the rate of return to education in Thailand for workers who were born between 1946 and 1967. This approach controls for unobservable individual characteristics, such as ability, that may bias the estimated rate of return to education. One unusual result is that there is a downward bias in the estimates of the rate of return to education based on individual data. This result holds for several disaggregations of the data by demographic characteristics, ruling out the aggregation bias explanation. Alternatively, the downward bias is explained by an opportunity cost argument. Individuals with greater ability have high potential wages, and therefore enter the labor force rather than continuing their education. This would imply a negative correlation between ability and years of schooling, and with a positive correlation between ability and earnings the individual data regressions would show a negative bias due to the omitted ability factor.

Based on the pseudo-panel estimations, the overall rate of return to education in Thailand is between 14% and 16%, which is considerably higher than estimated in prior studies that have used individual data from Thailand. Additional findings are that returns to education are higher

for females than for males, and unmarried individuals show higher returns than married workers. Not surprisingly, urban workers receive higher returns to education than rural workers due to their greater opportunities to exploit their increased skills in the cities.

The comparatively high rate of return to education found here, together with the opportunity cost argument behind the downward bias in the individual data regressions, leads to a policy recommendation. According to this analysis high ability individuals leave school early because the opportunity costs plus direct costs of education exceed the gains of additional schooling. Increased government subsidies of education could lower the direct costs to induce more high ability individuals to continue their education, raising their level of skills and productivity.

References

Amornthum, S. and Y. Chalamwong 2001. "Rate of Return to Education". *Human Resources and the Labor Market of Thailand*. Thailand Development Research Institute (TDRI)

Angrist, J. D. and A. B. Krueger .1991. "Does compulsory school attendance affect schooling and earnings?". *Quarterly Journal of Economics* 106(4): 979-1014.

Antman, F. and D. Mckenzie. 2007. "Poverty traps and Nonlinear Income Dynamics with Measurement Error and Individual Heterogeneity." *Journal of Development Studies*. October, 2007

Ashenfelter, O., C. Harmon., and O. Hessel. 1999. "A Review of Estimates of the Schooling/Earnings Relationship, with tests for Publication Bias". *Labor Economics*. 6(4): 453-70.

Becker, G.S. 1964. *Human Capital: A Theoretical and Empirical Analysis*. New York: National Bureau of Economic Research.

Becker, G. S. and B. R. Chiswick .1966. "Education and the distribution of earnings". *The American Economic Review*. 56(1/2): 358-369.

Belman, D. and J. Heywood.1991. "Sheepskin Effects in the Return to Education Examination of Women and Minorities". *Review of Economics and Statistics*. 73:720- 724

Bennell, P. 1996. "Rates of Return to Education: Does the Conventional Pattern Prevail in sub-Saharan Africa?" *World Development* 24(1): 183-199.

Behrman, J. and A. Deolalikar. 1995. "Are there differential returns to educations to schooling by gender? The case of Indonesian labour market". *Oxford Bulletin of Economics and Statistics*. 57(1): 97-117

Borjas, G. 1980. "The relationship between wages and weekly hours of work: the role of division bias". *Journal of Human Resources*. 15 (3): 409-423

Borton, J. 2003. "Thailand's novel education policy". Asia Times. (http://www.atimes.com)

Boockmann, B. and V. Steiner. 2006. "Cohort effects and the returns to education in West Germany".*Applied Economics.* 38: 1135-1152

Bound, J., D. Jaeger., and R. Baker. 1995. "Problems with Instrumental Variables Estimation

When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak". *Journal of the American Statistical Association*. 90(430): 443-50

Card, D. 1999. ''The Causal Effect of Education on Earnings,'' in *Handbook of Labor Economics*, Volume 3A, ed. by Orley Ashenfelter and David Card. Amsterdam and New York: North Holland.

Card, D. 2001. "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica* 69(5): 1127-1160.

Card, D. and Lemieux, T. 2001. "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis". *The Quarterly Journal of Economics*, 116 (2): 705-746

Chiswick, C. 1976. "On estimating earning functions for LDCs". *Journal of Development Economics*. 3:67-78

Chalongphob, S. and C. Yongyuth. 1996. Thailand development strategies and their impacts on labour markets and migration. In D. O'Connor, & L. Farsakh, *Development strategy, employment, and migration*. France

Deaton, A. 1985. "Panel data from a time series of cross-sections". *Journal of Econometrics*. 30: 109-126

Deaton, A. 1998. *The Analysis of Household Surveys*. The Johns Hopkin Univesity Press. USA

Devereux, P. 2006. "Improved errors-in-variables estimators for grouped data". *Journal of Business and Economic Statistics.* WP06/02

Devereux, P. 2007. "Small-sample bias in synthetic cohort models of labor supply". *Journal of applied econometrics.* 22: 839-848

Dougherty, C. 2005. "Why Are the Returns to Schooling Higher for Women than for Men?". *Journal of Human Resources.* XL(4): 969-988

Gandhi-Kingdon, G. 2002. "The gender gap in educational attainment in India: how much can be explained?". *The Journal of Development Studies*. 39(2): 25-53

Glenn, N. 2005. *Cohort Analysis*. Sage Publications (CA)

Grilliches, Z. 1977. "Estimating the Returns to Schooling: Some Econometric Problems". *Econometrica*. 45(1): 1-22

Harmon, C. and I. Walker. 1995. "Estimates of the Economic Return to Schooling for the United

Kingdom." *American Economic Review* 85(5): 1278-86.

Hawley, J.2004. “Changing returns to education in times of prosperity and crisis, Thailand 1985-1998”. *Economics of Education Review*. 23: 273-286

Heckman J., L. Lochner and T.Petra. 2005. “Earnings functions, Rate of return, and Treatment Effects: The Mincer Equation and Beyond”. *NBER Working Paper 11544.*

Kalwij, A. 2000. “Estimating the economic return to schooling on the basis of panel data”. *Applied Economics.* 32(1): 61-71

Kane, T. and C. Rouse. 1993. “Labor Market Returns to Two-and Four-Year Colleges: Is a Credit and Do Degree Matter?”. *NBER Working Papers: 4268*,1993

Mincer, J. 1974. *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research. 10

National Statistical Office of Thailand. 2007. Statistical Forecasting Bureau. *Labor Force Survey*, Thailand

Office of the National Education Commission (1997). *Education in Thailand 1997*. Bangkok. Thailand. Bureau of Educational System Development and Macro Planning.

Psacharopoulos, G. and H. A. Patrinos. 2004. “Returns to investment in education: A further update”. *Education Economics*. 12 (2): 111-134

The University of Chicago-UTCC research center at University of the Thai Chamber of Commerce. 2007. *Labor Force Survey 2005 Codebook*. March 2007. Thailand

Thosanguan, V. 1978. “The position of women and their contribution to the food processing industry in Thailand”. Workshop on TCDC and Women at Asian and Pacific Centre for women and development, Tehran, Iran (24-26 April, 1978)

UNDP. 2008. *Report on Thailand Gender Disaggregated Statistics.* UNDP, Thailand.

Verbeek, M. and J. Nijman. 1992. “Can cohort data be treated as genuine panel data?”. *Empirical Economics*. 17: 9-23

Verbeek, M. and T. Nijman. 1993. “Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections”. *Journal of Econometrics*. 59: 125-136

Welsh, F. 1997. “Wage and Participation”. *Journal of Labour Economics.* 15(1): 77-103

Table 1: Returns to education estimates for individual data, one-year cohort means, and two-year cohort means

	Individual Data (Cross-sectional regression) (1)	Pseudo-Panel (One-year cohort means) (2)	Pseudo-Panel (One-year cohort means) (3)	Pseudo-Panel (Two –year cohort means) (4)	Pseudo-Panel (Two- year cohort means) (5)
Constant	-0.0735	-0.389	-0.403	-0.418	-0.167
	(0.026)	(0.0447)	(0.0523)	(0.0551)	(0.816)
Years of education	0.115	0.145	0.151	0.149	0.161
	(0.00248)	(0.00417)	(0.00473)	(0.00547)	(0.00625)
Age	0.0838	0.0829	0.0802	0.0821	0.0636
	(0.00141)	(0.00212)	(0.00212)	(0.00255)	(0.0415)
Age squared	-0.000492	-0.000463	-0.000436	-0.000449	-0.000418
	(0.0000178)	(0.0000285)	(0.0000271)	(0.0000345)	(0.0000311)
Year dummies	Yes	Yes	Yes	Yes	Yes
Cohort dummies	-	No	Yes	No	Yes
Individual observations	199,833	199,833	199,833	199,833	199,833
Cohort-year observations	-	440	440	220	220
Individual observations per cohort					
- Max	-	1,017	1,017	1,690	1,690
- Min	-	113	113	284	284
Adjusted R^2	0.591	0.990	0.991	0.993	0.994

**standard errors are in parentheses*

Table 2: Returns to education estimates for men and women

	Men Individual Data (Cross-sectional regression) (1)	Men Pseudo-Panel (Two-year cohort means) (2)	Men Pseudo-Panel (Two-year cohort means) (3)	Women Individual Data (Cross-sectional regression) (4)	Women Pseudo-Panel (Two- year cohort means) (5)	Women Pseudo-Panel (Two - year cohort means) (6)
Constant	0.218	0.0279	-0.0870	-0.297	-0.601	-0.384
	(0.0369)	(0.0639)	(0.947)	(0.0353)	(0.0657)	(0.930)
Years of education	0.107	0.120	0.128	0.129	0.167	0.164
	(0.000340)	(0.00648)	(0.00747)	(0.000359)	(0.00610)	(0.00652)
Age	0.0778	0.0806	0.0829	0.0828	0.0765	0.0670
	(0.00192)	(0.00275)	(0.0483)	(0.00205)	(0.00322)	(0.0472)
Age squared	-0.000439	-0.000467	-0.000447	-0.000483	-0.000402	-0.000352
	(0.0000241)	(0.0000376)	(0.0000361)	(0.0000261)	(0.0000433)	(0.0000388)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Cohort dummies	-	No	Yes	-	No	Yes
Individual observations	112,419	112,419	112,419	87,414	87,414	87,414
Cohort-year observations	-	220	220	-	220	220
Adjusted R^2	0.548	0.990	0.991	0.663	0.990	0.992

**standard errors are in parentheses*

Table 3: Returns to education estimates for urban and rural residents

	Urban Individual Data (Cross-sectional regression) (1)	Urban Pseudo-Panel (Two-year cohort means) (2)	Urban Pseudo-Panel (Two-year cohort means) (3)	Rural Individual Data (Cross-sectional regression) (4)	Rural Pseudo-Panel (Two-year cohort means) (5)	Rural Pseudo-Panel (Two-year cohort means) (6)
Constant	-0.117	-0.431	0.257	-0.0898	-0.333	-1.655
	(0.0314)	(0.0580)	(0.807)	(0.0514)	(0.0681)	(1.139)
Years of education	0.115	0.146	0.158	0.114	0.141	0.136
	(0.000306)	(0.00601)	(0.00627)	(0.000430)	(0.00578)	(0.00711)
Age	0.0853	0.0841	0.0433	0.0875	0.0834	0.153
	(0.00166)	(0.00268)	(0.0412)	(0.00273)	(0.00368)	(0.058)
Age squared	-0.000488	-0.000460	-0.000438	-0.000596	-0.000515	-0.000483
	(0.0000208)	(0.0000358)	(0.0000327)	(0.000035)	(0.0000505)	(0.0000481)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Cohort dummies	No	No	Yes	No	No	Yes
Individual observations	135,248	135,248	135,248	64,585	64,585	64,585
Cohort-year observations	-	220	220	-	220	220
Adjusted R^2	0.607	0.993	0.994	0.574	0.980	0.983

**standard errors are in parentheses*

Table 4: Returns to education estimates for married and unmarried workers

	Non-Married Individual Data (Cross-sectional regression) (1)	**Non-Married Pseudo-Panel (Two-year cohort means) (2)**	**Non-Married Pseudo-Panel (Two-year cohort means) (3)**	**Married Individual Data (Cross-sectional regression) (4)**	**Married Pseudo-Panel (Two-year cohort means) (5)**	**Married Pseudo-Panel (Two-year cohort means) (6)**
Constant	0.0379	-0.199	1.005	0.270	0.257	0.462
	(0.0458)	(0.0813)	(1.048)	(0.0368)	(0.0466)	(0.867)
Years of education	0.126	0.152	0.155	0.112	0.119	0.107
	(0.000504)	(0.00612)	(0.00652)	(0.000282)	(0.00580)	(0.00692)
Age	0.0690	0.0653	0.00270	0.0713	0.0693	0.0649
	(0.00261)	(0.00389)	(0.0533)	(0.00185)	(0.00336)	(0.0440)
Age squared	-0.000377	-0.000304	-0.000281	-0.000342	-0.000325	-0.000324
	(0.0000342)	(0.0000523)	(0.0000519)	(0.0000228)	(0.0000441)	(0.0000415)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Cohort dummies	No	No	Yes	No	No	Yes
Individual observations	50,977	50,977	50,977	148,856	148,856	148,856
Cohort-year observations	-	220	220	-	220	220
Adjusted R^2	0.627	0.976	0.990	0.582	0.992	0.994

**standard errors are in parentheses*

Appendix A: Numbers of observations per cell: single-year cohort

YEAR	COHORT (OR BIRTH YEAR) 1967	1966	1965	1964	1963	1962	Total
1986	161	223	167	223	288	333	5,605
1987	170	164	203	229	245	331	5,065
1988	159	212	247	264	313	326	5,320
1989	339	353	359	435	375	417	7,081
1990	351	334	457	399	384	395	6,824
1991	431	506	448	454	476	410	7,920
1992	531	459	440	507	349	533	7,616
1993	451	431	464	405	494	330	7,187
1994	832	793	760	930	704	839	14,313
1995	931	714	1,017	747	864	732	14,394
1996	737	941	673	769	663	706	13,216
1997	770	519	651	592	502	663	10,011
1998	644	674	679	614	762	739	12,187
1999	763	613	640	819	756	669	11,993
2000	686	561	728	678	673	734	11,384
2001	728	802	751	711	814	672	12,593
2002	933	728	714	801	654	967	12,395
2003	776	693	740	720	835	652	11,924
2004	670	733	638	842	598	720	10,939
2005	801	681	924	695	795	747	11,866
Total	11,864	11,134	11,700	11,834	11,544	11,915	199,833

YEAR	COHORT (OR BIRTH YEAR) 1961	1960	1959	1958	1957	1956	Total
1986	392	318	307	348	334	402	5,605
1987	316	269	310	243	380	243	5,065
1988	331	340	279	365	243	281	5,320
1989	421	342	506	287	443	341	7,081
1990	353	474	290	332	396	310	6,824
1991	533	400	411	369	437	374	7,920
1992	389	426	357	317	470	357	7,616
1993	449	379	345	408	420	307	7,187
1994	734	681	865	764	807	796	14,313
1995	715	869	796	679	944	629	14,394
1996	864	743	697	661	720	779	13,216
1997	598	523	576	440	673	451	10,011
1998	655	751	644	846	636	646	12,187
1999	796	576	798	535	817	575	11,993
2000	614	762	513	660	715	499	11,384
2001	845	589	824	660	719	703	12,593
2002	587	802	670	508	900	508	12,395
2003	744	729	653	650	735	509	11,924
2004	661	593	774	507	688	547	10,939
2005	623	814	684	608	756	581	11,866
Total	11,620	11,380	11,299	10,187	12,233	9,838	199,833

YEAR	COHORT (OR BIRTH YEAR) 1955	1954	1953	1952	1951	1950	Total
1986	250	277	238	197	264	208	5,605
1987	254	262	182	249	228	190	5,065
1988	249	217	246	241	176	201	5,320
1989	305	344	309	264	259	192	7,081
1990	319	323	280	285	218	259	6,824
1991	412	304	333	275	326	193	7,920
1992	345	333	231	342	215	239	7,616
1993	358	266	318	221	234	194	7,187
1994	639	714	482	647	438	381	14,313
1995	841	491	612	528	410	509	14,394
1996	570	609	495	405	460	385	13,216
1997	513	370	329	409	285	269	10,011
1998	531	458	507	454	346	412	12,187
1999	477	495	466	384	393	316	11,993
2000	592	424	422	482	286	328	11,384
2001	546	431	544	409	394	290	12,593
2002	518	599	435	482	260	335	12,395
2003	607	461	532	321	358	302	11,924
2004	458	503	329	384	291	237	10,939
2005	638	350	409	377	298	318	11,866
Total	9,422	8,231	7,699	7,356	6,139	5,758	199,833

YEAR	COHORT (OR BIRTH YEAR) 1949	1948	1947	1946	Total
1986	161	186	159	169	5,605
1987	189	122	173	113	5,065
1988	147	199	145	139	5,320
1989	279	140	224	147	7,081
1990	146	210	191	118	6,824
1991	273	173	183	199	7,920
1992	215	174	246	141	7,616
1993	183	215	165	150	7,187
1994	470	343	330	364	14,313
1995	378	310	390	288	14,394
1996	356	373	288	322	13,216
1997	280	191	246	161	10,011
1998	297	335	256	301	12,187
1999	351	237	286	231	11,993
2000	262	286	271	208	11,384
2001	359	305	289	208	12,593
2002	311	253	250	180	12,395
2003	254	266	215	172	11,924
2004	253	176	170	167	10,939
2005	251	184	180	152	11,866
Total	5,415	4,678	4,657	3,930	199,833

Appendix B: Numbers of observations: two-year cohort case

YEAR	COHORT (OR BIRTH YEAR) 1966-1967	1964-1965	1962-1963	1960-1961	1958-1959	1956-1957	Total
1986	384	390	621	710	655	736	5,605
1987	334	432	576	585	553	623	5,065
1988	371	511	639	671	644	524	5,320
1989	692	794	792	763	793	784	7,081
1990	685	856	779	827	622	706	6,824
1991	937	902	886	933	780	811	7,920
1992	990	947	882	815	674	827	7,616
1993	882	869	824	828	753	727	7,187
1994	1,625	1,690	1,543	1,415	1,629	1,603	14,313
1995	1,645	1,764	1,596	1,584	1,475	1,573	14,394
1996	1,678	1,442	1,369	1,607	1,358	1,499	13,216
1997	1,289	1,243	1,165	1,121	1,016	1,124	10,011
1998	1,318	1,293	1,501	1,406	1,490	1,282	12,187
1999	1,376	1,459	1,425	1,372	1,333	1,392	11,993
2000	1,247	1,406	1,407	1,376	1,173	1,214	11,384
2001	1,530	1,462	1,486	1,434	1,484	1,422	12,593
2002	1,661	1,515	1,621	1,389	1,178	1,408	12,395
2003	1,469	1,460	1,487	1,473	1,303	1,244	11,924
2004	1,403	1,480	1,318	1,254	1,281	1,235	10,939
2005	1,482	1,619	1,542	1,437	1,292	1,337	11,866
Total	22,998	23,534	23,459	23,000	21,486	22,071	199,833

YEAR	COHORT (OR BIRTH YEAR) 1954-1955	1952-1953	1950-1951	1948-1949	1946-1947	Total
1986	527	435	472	347	328	5,605
1987	516	431	418	311	286	5,065
1988	466	487	377	346	284	5,320
1989	649	573	451	419	371	7,081
1990	642	565	477	356	309	6,824
1991	716	608	519	446	382	7,920
1992	678	573	454	389	387	7,616
1993	624	539	428	398	315	7,187
1994	1,353	1,129	819	813	694	14,313
1995	1,332	1,140	919	688	678	14,394
1996	1,179	900	845	729	610	13,216
1997	883	738	554	471	407	10,011
1998	989	961	758	632	557	12,187
1999	972	850	709	588	517	11,993
2000	1,016	904	614	548	479	11,384
2001	977	953	684	664	497	12,593
2002	1,117	917	595	564	430	12,395
2003	1,068	853	660	520	387	11,924
2004	961	713	528	429	337	10,939
2005	988	786	616	435	332	11,866
Total	17,653	15,055	11,897	10,093	8,587	199,833