ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

RETURNS TO EDUCATION IN SIR LANKA: A PSEUDO PANEL APPROACH

Rozana Himaz and Harsha Aturupane

Number 615
July 2012

Manor Road Building, Oxford OX1 3UQ

Title Returns to Education in Sri Lanka: A Pseudo Panel Approach

Authors Rozana Himaz (University of Oxford)

Harsha Aturupane (The World Bank)

Short Title: Returns to Education in Sri Lanka

Corresponding author: Rozana Himaz (corresponding author)

Queen's College, University of Oxford, Queen's College, OX1 4AW, UK

e-mail: rozana.himaz@queens.ox.ac.uk

Tel: +44 (1865) 281800
Fax: +44 (1865) 281801

Abstract

This study employs the pseudo-panel approach to estimate returns to education among income earners in Sri Lanka. Pseudo-panel data are constructed from nine repeated cross-sections of Sri Lanka's Labor Force Survey data from 1997-2008 for workers born during 1953–1974. The results show that for males, one extra year of education increases monthly earnings by about 5 per cent using the pseudo panel estimation rather than 8 per cent as in the OLS estimation. This indicates that not controlling for unobservables such as ability and motivation bias the OLS estimation of returns upwards by about 3 per cent on average. It also suggests that males with higher ability seem to be acquiring more years of education contrary to what has been observed recently in countries such as Thailand (Warunsiri and McKnown 2010) where the opportunity cost of education seems to be high.

Keywords: Sri Lanka, Education, returns, pseudo panels, synthetic cohorts

JEL Classification codes : I00 I20 I25 C23

1. Introduction

The estimation of returns to education both at the private and social levels has been central to the economics of human capital since the early 1960s, and has been the subject of much debate and discussion. The most common method used to analyse private returns to education has been based on the Mincer regression of log earnings on years of schooling and years of post school work experience. The basic model assumes that the log of earnings is a linear function of the level of education and experience and has been widely estimated (Psacharapolous 1981, Psacharapolous and Patrinos 2004, 2007, Heckman et. al. 2008). The coefficient on the schooling variable is often interpreted as an estimate of the internal rate of return. A key problem with the Mincer analysis is that it does not account for the endogeneity of the schooling variable. If unobservables such as ability and motivation are correlated to both schooling and wages, this may result in a bias in the coefficient of the schooling variable that is used to estimate private returns to education. The bias is likely to be positive if more able students are the ones who pursue more years of education. However, it could also be negative if, for example, the more able and motivated leave school early if presented with higher wage options.

The issue of unobservables such as ability and motivation causing biases in estimated returns has been the preoccupation of the literature since the earliest contributions in this area and a number of approaches have been used to deal with the issue. In some studies, measures of ability (e.g., intelligent quotient or IQ scores) are incorporated directly into the Mincerian wage equation to proxy ability. Doing so confirms that OLS estimates are biased upwards in studies such as Blackburn and Neumark (1993). More recent non-experimental methods have included instrumental methods, matching methods, and control function methods (Blundell et. al. 2004)[1]. For example Angrist and Krueger (1991), Angrist (1990), Card (1995), Harmon and Walker (1995) all attempt instrument for schooling outcomes with variables that are orthogonal to ability. Instruments used include quarter of birth, distance to school, the presence of a university or teacher

[1] Some studies such as Murphy and Welch (1990) and Katz and Autor (1999) argue that the functional form of the Mincer model itself is no longer applicable in some cases. Their example is that of U.S. workers. This begs the use of non-parametric methods to estimate returns. Heckman, Lochner and Todd find that non-parametric estimations of returns that account for tuition costs, income taxes and non-linearities in the earning-schoolings function lead to much higher results than those based on parametric estimations of marginal rates of return for some levels of schooling in the U.S. The non-linearity of the wage-schooling relationship is not yet an issue for Sri Lanka, as basic scatter plots indicate approximate linearity. Therefore this paper will only focus on estimating returns to schooling parametrically.

training college in the region of residence. Often, however, the use of instruments yields returns that are higher than the OLS estimations. This is probably due to the instruments being weak even though they may be valid and relevant, leading to estimators that perform poorly (Bound, Jaeger and Baker 1995, Staiger and Stock 1997).

Some studies have used panel data to account for the issue of unobservable fixed effects that might bias estimates. If genuine panel data was available, individual specific time invariant fixed effects can be accounted for by including individual history in the model, constructing instruments, by transforming models to first differences or by obtaining deviations from individual means. Although appealing, suitable panel data at the individual level relating to labour market and demographic characteristics are particularly hard to find for developing countries for a suitable length of time. Moreover, panel data often suffers from issues of non-response attrition and represent a small part of the labour force due to costs involved in data collection. In the absence of individual panel data, a modification of the more general fixed effects framework is applied by exploiting within-twins or within-siblings differences in wages and education if one were prepared to accept the assumption that unobserved effects are additive and common within twins (or siblings) so that they can be differenced out by regressing the wage difference within twins against the education difference (Ashenfelter and Zimmerman 1997). A central problem with such analysis is that unobserved effects may have an individual component as well as a family component which is not independent of the schooling variable. Thus although the family component is controlled, the individual component may not be, leading to results that may not be any less biased than those of ordinary least squares estimation. Another issue is that any error in measurement of the schooling variable will account for a larger fraction of differences between siblings than across the population as a whole. Forming differences between siblings will increase the bias from measurement error, causing a downward bias in estimates. One way to deal with this issue, is to adjust estimates using independent measures of error variance. Many within twin studies that control for measurement error suggest an ability bias that is relatively small.

In the absence of large panel datasets large cross section data sets repeated over time can be used to create pseudo panels or synthetic cohorts as suggested by Deaton (1985) to estimate returns under certain assumptions that control for unobserved individual specific effects including those such as ability and motivation. Pseudo panels are typically constructed from a time series of independent surveys conducted under the same methodology on the same reference population but in different time periods such as labour force and household surveys that can be found in many developing countries. The pseudo-panel is created by grouping individuals into criteria that do not change from one survey to another such as the year of birth or education level of the adult, assuming there are very limited options for changing the level of formal education in adulthood.

Much of the previous work on estimating returns to education in Sri Lanka has been based on estimating returns using the Mincer model and cross section data (Gutkind 1984; Glewwe 1985; Sahn and Alderman 1988; Aturupane 1993). The estimations of returns overall of around 15 per cent for secondary education (i.e., completed 8 to 13 years of schooling). Later papers by Himaz (2010) and Gunawardane (2002) use a fixed effects method based on cross section data to look for within-sibling differences in wage returns . These results indicate that OLS estimations based on the Mincer model are upward biased with average returns for the secondary education category being only around 6 percent. However, as only sibling data is used, only a sub-sample of many of the data sets are used, especially as the those siblings still living in the same household tend to be those who are younger workers. Moreover the attempt is not completely effective in controlling for unobservable fixed effects if they are at the individual level. Few other recent papers including Himaz and Aturupane (2011), De Silva (2009) have looked at returns using quantile regression analysis to look at returns for various income groups. Although informative if schooling is considered exogenous, the results may be biased if unobservables such as ability and motivation are correlated to both the schooling variables and the error term.

In this paper, we attempt to estimate returns to education using pseudo panels, as an alternative to the Mincerian approach that uses cross-section data. Under this method, a series of 9 repeated cross sections taken from the Sri Lanka Labour Force Surveys for 1997-2001, 2003-2004, 2006 and 2008, are used to construct a panel data set based on cohort means for those born between 1953 to 1974. This makes the youngest worker 24 years of age and the oldest 55 years of age in the sample. We also present estimations for returns based on cross-section data using the Mincer model for comparison. As a check of robustness and comparison, we also estimate a fixed effects model of labour returns based on within sibling data.

The analysis is an important first step towards utilising the many good quality, rich cross section data sets available in Sri Lanka to analyse trends of a longitudinal nature using pseudo panel techniques. If results indicate that ability is a significant variable in acquiring more education, this has implications for both education and labour market policy. The method used can be extended to other developing countries that have good repeated cross sections of data but no true panel data to estimate returns.

The rest of the paper is organised as follows: The next section looks at the conceptual framework and empirical specification. Section 3 discussed the data. Section 4 looks at the results from pseudo panel analysis and compares it with results from OLS and sibling-based fixed effects analysis. Section 5 concludes.

2. Conceptual framework and empirical specification

The standard Mincerian earnings function (Mincer 1974) is specified as follows:

$$\mathrm{w} = \alpha + \beta_0 s + \beta_1 x + \beta_2 x^2 + \varepsilon \qquad (1)$$

Where w is the natural logarithm of earnings, s the years of schooling and x is years of work experience often proxied by age. This equation can be specified for time *t* (with *t=1..T*) and individual *i* (with *i= 1....N*) as follows:

$$\mathrm{w}_{\mathrm{it}} = \gamma + \beta_0 s_{it} + \beta_1 x_{it} + \beta_2 x_{it}^2 + \alpha_{it+} \varepsilon_{it} \qquad (2)$$

where *w* is the log of earnings earnings for individual *i* at time *t*, *s* is the number of years of schooling for individual *i* at time *t* and *x* is the number of years of experience (age) for individual *i* at time *t*. The term α_{it} captures unobserved individual heterogeneity that includes ability and motivation. If this was uncorrelated with the explanatory variables s_{it} and x_{it} then the model in (2) can be estimated consistently from cross section data using ordinary least squares treating $\alpha_{it+} \varepsilon_{it}$ as the composite error term. However, it is likely that α_{it} is correlated to both schooling and experience (s_{it} and x_{it} respectively). If α_{it} is observable, it can be included directly into the equation. However, in the absence of such information, unobservables represented by α_{it} will cause a least squares estimation to be biased. If genuine panel data was available, then individual specific fixed effects can be accounted for by including individual history in the model, constructing instrument or by transforming models to first differences or by obtaining deviations from individual means. However, in the absence of panel data, Deaton (1985) suggests that cohorts constructed from repeated cross section data can be used to estimate a fixed effects model. The cohorts, *c,* are defined by a shared common characteristic, such that each individual is a member or only one cohort. In our case, this is the year of birth. If all observations in the cohorts are aggregated, the resulting model can be written as:

$$\overline{\mathrm{w}_{\mathrm{ct}}} = \beta_0 \overline{s_{ct}} + \beta_1 \overline{x_{ct}} + \beta_2 \overline{x_{ct}^2} + \bar{\alpha}_{ct+} \bar{\varepsilon}_{ct} \qquad (3)$$

Where c=1,...C and t=1,...T and w_{ct} is the average of all monthly earnings for all individuals in cohort *c* at time *t* and similarly for the other variables in the model. If cohorts are defined by those born between 1953 and 1974, this gives us a pseudo panel of 21 cohorts over 9 time periods based on 9 cross section surveys from 1997 to 2008. Estimating β_0 from (3) can be still problematic, however, as $\bar{\alpha}_{ct}$ depends on t, and is likely to be correlated to x_{ct} as α_{it} was likely to be correlated to x_{it}. As $\bar{\alpha}_{ct}$ is unobservable it cannot be included directly in the estimation. However, $\bar{\alpha}_{ct}$ can be treated as a fixed unknown parameter with $\bar{\alpha}_{ct} = \alpha_c$ over time, if there existed a sufficiently large number of individual observations in each cohort (Verbeek 2007: 5). In this case the model can be written as:

$$\overline{\mathrm{w}_{\mathrm{ct}}} = \gamma + \beta_0 \overline{s_{ct}} + \beta_1 \overline{x_{ct}} + \beta_2 \overline{x_{ct}^2} + \alpha_{c+} \bar{\varepsilon}_{ct} \qquad (4)$$

As Warunsiri and Mcnown (2010:1618) note, all error components in (2) that are correlated with the explanatory variables have been purged from the error term in (4). This makes the fixed effects estimation consistent.

The error term in (4) can be assumed normal, independent and homoskedastic if the cohort size is fixed over time. However, if the cohorts are very different in size, this will mean that the error term is heteroskadistic and needs to be corrected by weighting each observation with the square root of cohort size (Deaton 1985:117). As cohort size can vary in our data, we use weighted least squares estimation.

For the sake of comparison and to check for robustness of results, we also conduct a sibling-based fixed effects analysis on the same dataset. Under this method, unobserved heterogeneity at the household level is corrected for by using fixed effects estimation to a cluster sample, where the well-defined cluster is the household in the pooled data set. I use deviations from household means for all households where there are two or more males who are wage earners, assuming that differences are across households (if they exist). If there are unobservable fixed effects and they are significant, the constant term of the fixed effects regression would be significant and the OLS estimations would be biased (Pitt and Rosenzweig 1990). The fixed effects method addresses the issue of omitted variable bias arising from unobserved heterogeneity at the household levels only if such unmeasured attributes are common to individuals in the same household. If the unobserved household effects are random instead of being fixed, they would bias the error term and invalidate standard statistical tests. In order to test for this possibility, I estimate a random effects model using the same subsample. I then use the Hausman test to compare between the fixed and random effects estimations. The results are then compared to both the pooled OLS and pseudo panel estimates.

3. Data and descriptive statistics

The data comes from a series of 9 repeated cross sections of the Sri Lanka Labour Force Surveys for the years 1997-2001, 2003-2004, 2006 and 2008. The Sri Lanka Labour Force Survey is representative of the entire country apart from the Northern and Eastern provinces for which data was not available for most of the years. The Quarterly Labour Force Survey has been carried out since 1990 to produce estimates of employment, unemployment, labour force and basic demographic characteristics. The sampling method used is a stratified two stage sample design. In the first stage, the provinces of the country were divided into sectors. From these domains, census blocks were selected according to the population distribution (based on census data) so that the probability of selection is proportional to the size of the population in the block. In the second stage, ten households were selected from each block randomly. The entire sample was then divided randomly into four groups, one for each quarter. The sample includes all persons in the household aged 10 or above. A household or housing unit is defined as a house, apartment or room (s) that is occupied as a separate living space. The survey excludes housing units with 5 or more lodgers and institutions such as hospitals or military camps. Although the data has been collected quarterly in most years we have aggregated it by year for the purpose of this study. We use data from 1997 onwards mainly as some of the previous surveys are slightly different in design. Most notably, some of the previous surveys include only detailed information on earnings from paid employment (wages and salaries) rather than all earnings including self employment and own account work.

The nine repeated cross sections of data are used to construct a panel data set based on cohort means for those born between 1953 to 1974. This makes the youngest worker 24 years of age and the oldest 55 years of age in the sample. We also construct further disaggregated panels for males and workers from rural areas. We do not conduct similar panels for females and those in urban areas as the cohort sizes are very small (sometimes less than 100 observations each), and as discussed later, this may result in inaccurate standard errors in pseudo panel analyses.

Monthly earnings used are those from the primary employment, for those who usually work for more than 35 hours a week, which accounts for over 90 per cent of the men in the sample aged 24-55 who are employed. The earnings are in real terms, deflated by the GDP deflator with the base year being the year 2002. In all 9 survey years, roughly 86 per cent of the men are employed. Of those employed, 60 per cent are wage employees, 4 per cent are employers, and around 32 per cent are own account workers. A small percentage of around 3 per cent are classified as unpaid family workers. As we use earnings rather than wages as the dependent variable, we use information for almost all employed males and females born between 1953 and 1974 in the sample.

The level of schooling an individual has obtained is reported in terms of years of schooling completed, in the surveys. This ranges from 0 to 10 with 0 referring to having had no schooling at all, 1-5 years reflecting primary schooling, 6 to 13 secondary schooling, and over 14 years of education reflecting tertiary education received from universities and polytechnics.

We also use some information on some basic characteristics of the cohort such as age to capture experience, the square of age to capture any non-linearities with respect to age-related experience, ethnicity (the proportion of the cohort that is Tamil or belongs to a Minority group, with Sinhalese omitted) and sector of residence (the proportion that live in urban and estate areas with rural omitted).

Basic descriptive statistics for the entire sample, based on individual information is presented in Table 1 below. The first column indicates summary statistics for the full sample based on pooled data that comprises 54 759 males and females born between 1953-1974. The natural log of earnings is 8.43 (about Rs. 4582 a month). The average years of education in the sample is 8.5 and the average age is 38. Roughly 80 per cent of the sample contains the majority Sinhalese. Around 68 per cent live in rural areas, 21 per cent live in urban areas and 12 per cent in estate areas. The second column contains information for males only, who form over 60 per cent of the full sample. The earnings are slightly higher than for the full sample, suggesting that women earn less on average than men. The males only sample also indicates a slightly higher education.

Table 1: Descriptive statistics based on pooled individual level data

	Full sample-males and females born between 1953-1974	Sub sample-males born between 1953-1974
Monthly earnings (natural log)	8.438 (0.880)	8.523 (0.109)
Education in years	8.508 (3.947)	8.625 (0.426)
Age	38.227 (6.939)	38.778 (7.023)
Aged squared	1509.502 (538.47)	1552.778 (549.64)
Tamil	0.162 (0.368)	0.125 (0.029)
Muslim	0.045 (0.207)	0.057 (0.020)
Urban	0.214 (0.410)	0.226 (0.061)
Estate	0.119 (0.323)	0.083 (0.037)
Individual observations	54759	29279

4. Results

The results using the full sample are reported in Table 2 below. The first column show the results based on Ordinary Least Squares estimation based on individual cross section data. The estimation includes 8 dummy variables to capture yearly fixed effects. The next two columns show results based on pseudo panel data. The first of these is based on means for yearly birth cohorts from 1953 to 1974. The second is based on means for two year groups from 1953 to 1974. The analysis is conducted using two different groupings of the data to ensure results are robust.

Table 2: Returns to education estimates for individual data, one year cohort means, two year cohort means and sibling fixed effects

VARIABLES	Individual data (cross sectional regression)	Pseudo panel One year cohort means	Pseudo panel two year cohort means	Sibling Fixed Effects
Educ	0.0861***	0.0711***	0.0853**	0.044***
	(0.000959)	(0.0270)	(0.0360)	(0.0029)
Age	0.0343***	0.0442***	0.0422***	0.0001
	(0.00486)	(0.0116)	(0.0116)	(0.0001)
age2	-0.000386***	-0.000389***	-0.000426***	
	(6.25e-05)	(0.000143)	(0.000139)	
Tamil	0.00649	-0.379	-0.720	
	(0.0132)	(0.368)	(0.520)	
Muslim	-0.00710	-0.463	-0.226	
	(0.0166)	(0.509)	(0.744)	
Urban	0.230***	-0.264	-0.501**	
	(0.00871)	(0.200)	(0.242)	
Estate	-0.0614***	0.544**	0.697**	7.735***
	(0.0149)	(0.258)	(0.316)	(0.2259)
Constant	6.705***			
	(0.0926)			
Individual Observations	54759	54759	54759	18912
Cohort-year observations	-	189	99	-
Mean observations per cohort	-	318	608	-
Number of households with two or more members earning	-	-	-	8978
R-squared	0.191			
Hausman test H_0: Difference in coefficients not systematic				$\chi^2(3)=405.77$

Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Returns measured by OLS analysis indicates that an extra year of education increases income by around 8 per cent. The returns measured using the one year pseudo panel analysis is slightly lower at around 7 per cent. The last column of the table reproduces results for a sibling-based fixed effects analysis on the same dataset. The data for all years is pooled and only those households that have two or more members earning who were born between 1953 and 1974 are considered. This severely restrict sample size and only about 34 per cent of the households in the sample fall into this category. If we did not define the birth year of the individuals, around 50 per cent of the sample have two or more individuals earning. We report results only for those born between 1953 and 1974 as this is comparable to the sample we use for the pseudo-panel analysis, in spite of the sample being severely restricted. The fixed effects method addresses the issue of omitted variable bias arising from unobserved heterogeneity at the household levels only if such unmeasured attributes are common to individuals in the same household. If the unobserved household effects are random instead of being fixed, they would bias the error term and invalidate standard statistical tests. In order to test for this possibility, I estimate a random effects model using the same subsample. I then use the Hausman test to compare between the fixed and random effects estimations. The results are then compared to both the pooled OLS and pseudo panel estimates. The Hausman test, testing the hypothesis that difference in coefficients not systematic is rejected with a Chi-sqaured value of 26.29. Thus a fixed effects model is more appropriate. The estimations indicate a return to education that is much lower than that of the OLS or pseudo panel analysis. This result, however, should be interpreted with much caution as only 34 per cent of the sample have households with more than one member working who is also born between the years 1973 and 1953.

On the face of it, the results in Table 2 seem to suggest that using cross section data and pseudo panel data do not yield results that are very different. However, this is not the case as disaggregated analysis reveals a very different picture. Traditionally, returns are estimated separately for males and females. The labour force participation rates of males and females are quite different, as are the earnings and other labour market related characteristics. We can disaggregate and perform the analysis of returns only for males and not females. This is because females comprise only about 40 per cent of the sample, which leaves cell sizes too small for efficient standard errors is the pseudo panel analysis.

Table 3 produces results for a sub sample based on males only. Returns measured using OLS analysis indicates that an extra year of education increases income by 8 per cent. Quite notably, the returns measured using pseudo panel techniques reported in columns 2 and 3, are lower at around 5.5 per cent. This suggests that when unobservables such as ability and motivation biases upwards the returns to education in cross section analysis. In other words, males with higher ability seem to acquire more education in the Sri Lankan context. This result is in sharp contrast to what Warunsri and McOwen (2010) find for Thailand, where the returns measured using pseudo panel data are higher than that measured using OLS estimation suggesting that the opportunity cost of education is high for high ability workers. Their results suggest that higher ability workers leave formal education sooner than lower ability workers due to perhaps attractive wage rates in the labour market.

Our results also show that experience is an important contributor to earnings, with every year of experience adding to roughly 4 per cent increase in earnings in both the cross-section and pseudo panel results. The cross section results also show that the individual being of Tamil ethnicity exerts a significant negative impact on earnings compared to Sinhalese ethnicity and that residence in the estate sector exerts a significant negative influence compared to rural residence. Neither of these results are unambiguously confirmed in the pseudo panel results reported using one year and two year cohort means.

Table 3: Returns to education estimates for individual data, one year cohort means and two year cohort means for males

VARIABLES	Individual data (cross sectional regression)	Pseudo panel One year cohort means	Pseudo panel two year cohort means
Educ	0.0808***	0.0550**	0.0574*
	(0.00121)	(0.0258)	(0.0321)
Age	0.0352***	0.0455***	0.0428***
	(0.00594)	(0.0128)	(0.0123)
age2	-0.000395***	-0.000375**	-0.000401***
	(7.63e-05)	(0.000156)	(0.000146)
Tamil	-0.0501***	-0.544	-0.756*
	(0.0165)	(0.351)	(0.435)
Muslim	-0.0249	-0.472	-0.185
	(0.0181)	(0.397)	(0.527)
Urban	0.213***	-0.280	-0.564**
	(0.0105)	(0.188)	(0.219)
Estate	-0.163***	0.424	0.488
	(0.0194)	(0.285)	(0.315)
Constant	6.979***		
	(0.113)		
Observations	37574		
Cohort-year observations		189	99
Mean observations per cohort		230	440
R-squared	0.165		

Note: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

5. Conclusions

This paper is a first attempt at utilising a vast wealth of good quality repeated cross section data in Sri Lanka to build pseudo panel data sets for analysing complex socio-economic issues. It looked at returns to education in Sri Lanka based on conventional cross section data using the Mincerian equation, and compared it with pseudo panel data constructed using 9 Labour Force Surveys from 1997 to 2008. When men and women are pooled together, estimated returns do not differ very much between methods: OLS using cross section and controlling for fixed effects using pseudo panel data. However, when the analysis is disaggregated between men and women, it is clear that the OLS analysis has a notable ability bias that inflates the coefficient on the schooling variable: An extra year of education increases monthly earnings by about 8 per cent using the OLS estimation rather than 5 per cent as in the pseudo panel estimation. This is suggestive that not controlling for unobservables such as ability and motivation bias the OLS estimation of returns upwards by about 3 per cent on average for males. It also suggests that males with higher ability seem to be acquiring more years of education contrary to what has been observed recently in countries such as Thailand (Warunsiri and McKnown 2010) where the opportunity cost of education seems to be high. Unfortunately, we are not able to perform the same analysis for women and obtain reliable results as the cohort sizes become too small. Overall, the ability bias does seem to inflate returns estimated using OLS analysis and the Mincerian equation upwards, as is consistent with similar work done for countries such as the US.

References

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

O. Ashenfelter and D. Zimmerman, 1997 Estimates of the returns to schooling from sibling data:Fathers, sons and brothers', *Review of Economics and Statistics*, 79:1-9

Aturupane, H 1993 Is education beneficial? A microeconomic analysis of the impact of education on economic welfare in a developing country', Ph.D. diss., Cambridge University, UK

J. Bound, 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) (1995), pp. 443–450.

R. Blundell, L. Dearden and B. Sianesi,, 2001 Estimating the returns to education: Models, methods and results, Centre for the Economics of Education, London School of Economics, London

D. Card 1995 Using geographic variation in college proximity to estimate the return to schooling. In: L. Christofides, K. Grant and R. Swindinsky, Editors, *Aspects of labour market behaviour: Essays in honour of John Vanderkamp*, University of Toronto Press, Toronto

A. Deaton 1985, Panel data from a time series of cross-sections, *Journal of Econometrics* **30** (1985), pp. 109–126.

A. Deaton, 1998, The analysis of household surveys: A microeconometric approach to development policy, The Johns Hopkin Univesity Press, Baltimore, MD

I. De Silva, 2009Ethnicity and sheepskin effects in the returns to education in Sri Lanka: A conditional quantile analysis, *International Journal of Development Issues*, Emerald Group Publishing, vol. 8(1):61-79

C. Dougherty, 2005, Why are the returns to schooling higher for women than for men?, *Journal of Human Resources* **40** (4) (2005), pp. 969–988.

Glewwe, P (1985) An analysis of income distribution and labour markets in Sri Lanka, Ph.D. Diss, Stanford University

Z. Grilliches, 1977 Estimating the returns to schooling: Some econometric problems, *Econometrica* **45** (1) . 1–22.

E. Gutkind, 1984, Earnings functions and returns to education in Sri Lanka: Is education enough? *Sri Lanka Economic Journal*, vol 3(2): 57-103

Harmon, C., Oosterbeek, H., & Walker, I., 2000 . *The returns to education: a review of evidence, issues, and deficiencies in the literature*. Center for the Economics of Education, London School of Economics.

Heckman, J., Lochner, L., & Petra, T., 2005. *Earnings functions, rate of return, and treatment effects: The mincer equation and beyond.* NBER Working Paper: 11544.

R. Himaz, 2010, Intrahousehold allocation of education expenditure in rural Sri Lanka, *Economic Development and Cultural Change,* 58:2: 137-159

R. Himaz, and H. Aturupane (2011) Education and Household Welfare in Sri Lanka from 1985 to 2006," *Economics Series Working Papers* 527, University of Oxford, Department of Economics.

J. Mincer, 1974 Schooling, experience, and earnings, National Bureau of Economic Research, New York.

H.A. Patrinos, 1995, Education and earnings differentials, Mimeo, World Bank, Washington, DC.

G. Psacharopoulos, 1994, Returns to investment in education: A global update, *World Development* **22** (9): 1325–1343.

M. Pitt and M. R. Rosenzweig 1990, Estimating the behavioural consequences of health in a family context: the intrafamily incidence of infant illness in Indonesia', *International Economic Review* 31:969-89

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

D. E. Sahn, . and H Alderman 1988 The effects of human capital on wages, and the determinants of labour supply in a developing country*, Journal of Development Economics* 29(2):157-183

M. Verbeek and T. Nijman, 1992, Can cohort data be treated as genuine panel data?, *Empirical Economics* **17**, pp. 9–23.

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

S.Warunsiri & R. McNown, 2010The Returns to Education in Thailand: A Pseudo-Panel Approach, *World Development*, Elsevier, vol. 38(11):1616-1625