CREDIT Research Paper

No. 18/08

Rates of return to education in Sri Lanka

by

Nirodha Bandara, Simon Appleton and Trudy Owens

Abstract

This paper explores the rates of return to education in Sri Lanka across the sexes and different types of employment during 2009/10. The endogeneity bias suggests that education may be associated with other characteristics such as ability and family background – excluding such attributes could lead to biased estimates of the return to education. To deal with this form of bias, the paper uses an Instrumental Variables (IV) approach to measure education. Our results reveal that the OLS estimates are downward biased, in comparison to the IV estimates. Further, we add to the literature by exploring the return to education across levels of education – primary, secondary and tertiary – and how it varies across sexes and types of employment. Finally for the self-employed, we analyse the association between inputs (land, raw materials, labour and capital), education and output using a household-level production function – for those in agriculture and non-agricultural activities. The results suggest that education at higher levels have a positive association with non-agricultural output, but not with agricultural output.

JEL Classification: C26, I26, O12

Keywords: Instrumental variable estimation, Rate of return to education, Sri Lanka

Centre for Research in Economic Development and International Trade, University of Nottingham

CREDIT Research Paper

No. 18/08

Rates of return to education in Sri Lanka

by

Nirodha Bandara, Simon Appleton and Trudy Owens

The Authors

Nirodha Bandara is the Directress of Academic Affairs at Royal Institute of Colombo, Sri Lanka and the Corresponding Author (email: nirodha@royalinstitute.org). Simon Appleton is a Professor and Head of School of Economics, University of Nottingham Ningbo, China (simon.appleton@nottingham.edu.cn) and Trudy Owens is an Associate Professor of Economics, University of Nottingham, UK (email: trudy.owens@nottingham.ac.uk).

Acknowledgements

The authors are grateful to the comments and feedback given by Julie Litchfield (University of Sussex), and Sarah Bridges and researchers from the University of Nottingham, School of Economics.

Research Papers at www.nottingham.ac.uk/economics/credit/

1. Introduction

This paper examines the returns to education across the sexes and by different types of employment, and how returns vary at different levels of education (primary, secondary and tertiary) in Sri Lanka during 2009/10. The paper deals with the issue of endogeneity of education when estimating earnings, which has not been explored in the context of Sri Lanka[1]. For the self-employed, a household production function is estimated to identify the link between inputs, labour, land, education and other household characteristics, and agricultural and non-agricultural output/productivity. Evidence shows that individuals with higher education and more skills have an ability to earn more income. This is believed to have social benefits on groups of individuals, and hence benefit the country as a whole by means of higher GDP, poverty alleviation and democratic stability, for example. The theoretical framework on the effect of education on earnings based on Schultz (1961), Becker (1964) and Mincer (1974) has been studied extensively.

Several studies have explored the link between education and earnings, concluding that they share a positive relationship – more educated individuals have higher productivity levels and benefit from higher earnings compared to their less-educated counterparts (Psacharopoulos, 1994; Psacharopoulos and Patrinos, 2004; Kingdon et al., 2008). However, Arrow (1973) discusses that this link is valid if observed differences in earnings across individuals with different levels of education reflect true differences in productivity, rather than differences in ability. Ability is defined broadly as any unobservable factors (these could be non-cognitive, for example, family background) that affect earnings via education. Ignoring the effects on such unobservable characteristics could lead to an imprecise estimation of the returns to education.

Building on a large literature on the returns to education, this paper considers how returns vary across sectors of employment and by gender in Sri Lanka. The employment sectors include agricultural and non-agricultural workers, with a further sub-division of the non-agricultural workers to distinguish between private sector workers and those in the government sector. In order to measure the returns to education, an Instrumental Variables (IV) approach is used. This method allows us to deal with the potential endogeneity bias that could lead to an over or under-estimation of the returns to education. The endogeneity bias suggests that the level of education received by an individual may depend on unobservable characteristics such as ability and family background which should be taken into consideration. In order to do this, we make use of two instruments to get a better measure of education – spouse's level of education and a change in schooling laws in 1997 which led to an increase in the minimum years of compulsory schooling to 14 years. After testing for weak instruments and validity, the two instruments were used in the IV estimation.

1 To the best of our knowledge, based on a literature search in EconLit

In order to examine the variation in returns to education by level (primary, lower secondary, upper secondary and tertiary education) a corresponding OLS estimation will be presented. The final part of the analysis estimates a household production function for the self-employed agriculture and non-agriculture sectors. We will examine how the combination of inputs (raw materials, labour and capital) and education levels are associated with agricultural/non-agricultural output.

The IV estimation in comparison to OLS suggests that the latter produces returns to education that are downward biased. The returns to education are highest for private non-agricultural workers and women, even after controlling for occupation and industry. The rates of returns to education at various levels suggest that individuals working in the government sector receive the highest returns at lower levels of education (primary and lower secondary) whereas individuals working in the private sector receive higher returns to upper secondary and tertiary education.

By gender, the results suggest that men receive higher returns to education at lower levels whereas women receive higher returns to upper secondary and tertiary education. The findings from the production function suggest that inputs, such as raw materials, have a positive association with output. Education at the higher level is positively related to non-agricultural output; however, it is adversely related to agricultural output. Thus, more educated individuals benefit in non-agricultural self-employment. By contrast, more educated individuals working in agriculture appear to contribute less towards agricultural output.

The rest of the chapter will be organized as follows. Section 2 reviews the existing literature on the rates of return to education along with extensive motivation as to why Instrumental Variables was chosen as the main method of analysis. Section 3 describes the data, and includes an explanation of the methods being used. Section 4 outlines the results. The first part of this section compares OLS and IV, along with the tests to ensure that the instruments are valid. The second part of Section 4 reports an OLS disaggregation of the years of schooling to look at how the coefficient on the education variable changes across different levels of education. The final part of Section 4 presents an analysis of the production function for the private agricultural and non-agricultural self-employed enterprises/households. Lastly, Section 5 concludes with suggestions for future research.

2. Literature review

2.1 Returns to education: Theory and macroeconomic evidence

Some see education as a basic need; with equal access helping to fulfil other basic needs like sanitation, health and shelter. Education plays a vital role in helping people stay out of income poverty. As Appleton (2001) notes, education gives individuals better access to employment. Reversely, a person who does not have sufficient earned income would find

it difficult to attain more education; households are less likely to go back and get more education once they start earning. Several studies of human capital have explained that schooling can increase earnings and productivity levels, via more skills and knowledge (Becker, 1994).

The extent to which education increases earnings is known as the economic "return" to education. Investing in education brings about both, private and social returns. The private returns explain individual behaviour in their choice of education level and type. The social returns can be used to understand which area of education requires further investment; education can cause a spill over effect to other individuals, and as such, social benefits are realised. Colclough et al. (2010) state that the social rates of return to education are lower than the private rates of return to education possibly due to the addition of "publicly financed costs" to the private returns.

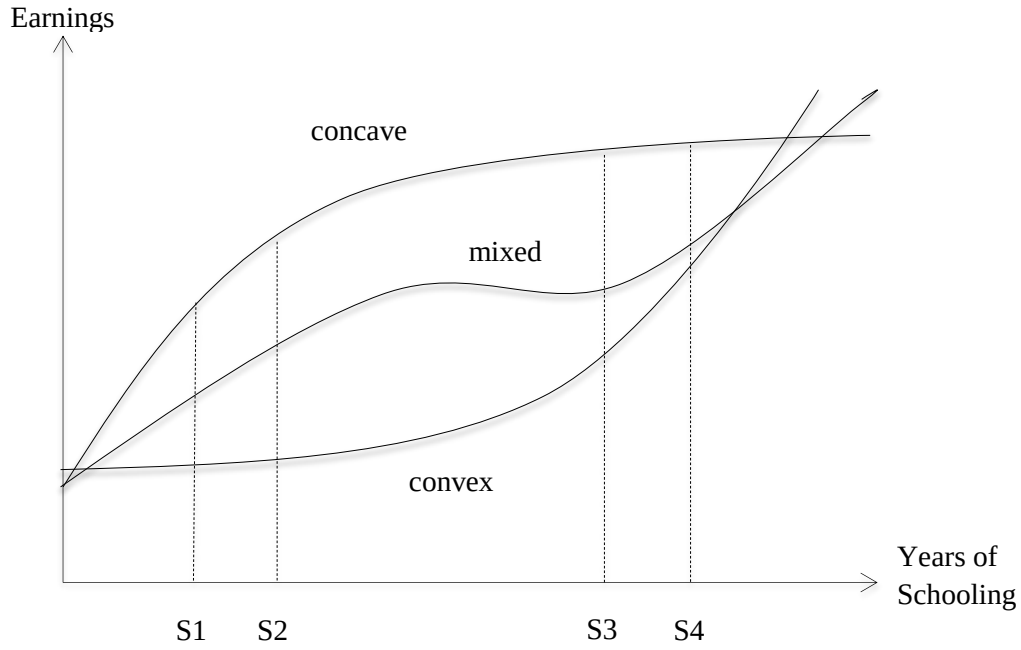
The entire economic impact of education may not be captured by pecuniary measurement - the benefits of education can be both monetary and non-monetary, the latter being difficult to measure (Vila, 2000). Wages are the monetary private returns, while non-monetary private returns include health benefits, enhancement of children's education and non-monetary job satisfaction. Monetary social benefits include positive effects on the country's Gross Domestic Product and the impact it has on others' earnings (by encouraging them to boost their levels of productivity). Non-monetary social benefits include for instance, poverty reduction, democratic stability and low crime rates.

This chapter focuses on the monetary private returns. It goes beyond an emphasis on wage income, and considers other forms of earned income. In a developing country such as Sri Lanka, a large proportion of the population earn self-employment income from agriculture, construction, trade, and so on. Colclough et al. (2010), in their analysis of several studies that look at the returns to education, state that wage earners are a minority of the working population in most developing countries, and that the returns to education in self-employment, especially in industries such as agriculture, may be different from traditional findings. Datt and Gunerwardena (1997) state that nearly half of those trapped by poverty depend on agricultural income; and another 30% depend on non-agricultural rural activities.

There is a common belief that education pays off through wage employment substantially more, relative to agricultural employment (Bennell, 1996; Appleton, 2001). On the other hand, both, cognitive and non-cognitive effects of education may have an impact on agricultural productivity (Appleton and Balihuta, 1996). Cognitive effects include obtaining certain skills such as literacy and numeracy, while non-cognitive effects work through the change of peoples' attitudes and practices – for example, making them aware of the possibility of improving their standard of living. Thus, the wage and self-employed are included in our analysis.

The literature considers the returns to education at various levels –primary, secondary and tertiary education. Psacharopoulos (1994) and Psacharopoulos and Patrinos (2004) suggest that primary education has higher returns than tertiary education partly because of the fact that tertiary education is more expensive relative to primary education. This concave relationship between earnings and education can be seen in Figure 1. Recent evidence, however, has shown that the rates of return to post-primary education have increased – thus the returns are increasing with the level of education, rather than decreasing (Colclough et al., 2010). In such a case, an extra year of education brings about a higher impact on earnings at post-primary education, compared to primary. This is shown by the convex curve in Figure 1 – the slope of the curve and the returns to education are increasing when moving to higher levels of education.

Figure 1: Education-Earnings relationships



Source: Colclough et al. (2010)

However it is important to note that the rates of return to education discussed by Psacharopoulos (1994) are defined as the value of education relative to its cost. Psacharopoulos, along with several other preceding studies made use of the Mincer earnings function $[log(wage) = \beta_o years\_of\_schooling + \beta_1 age + \beta_2 age^2]$ developed by Mincer (1974). This function identifies the relationship between wages and education, age and age-squared, while the parameters β_0, β_1 and β_2 yield the respective returns. The reason for the inclusion of age is because it is believed to be a proxy for experience (learning on the job). Many studies include additional variables such as occupation, gender and marital status to get a better estimate of the return to education. In the function given above, the coefficient on "years of schooling" is interpreted as the wage premium –

an additional year of education increases wages by say, 10 per cent – this is known as the Mincerian return to education[2].

The micro-level literature generally shows a 10 per cent return to education – an extra year of schooling brings about a 10 per-cent increase in wages (Psacharopoulos and Patrinos, 2004). The macro-level evidence, in most cases, shows that countries with a more educated workforce tend to benefit from higher GDP growth. The link is thought to be through higher productivity levels due to education, which in turn affects national income. Education may also affect creativity and innovation, and in turn, affect growth. Evidence suggests that countries with literacy scores one per-cent higher than the international average have labour productivity and GDP per capita 2.5 and 1.5 per cent higher respectively than other countries (OECD, Education at a Glance, 2006). Krueger and Lindahl (2001) found that a change *and* the initial level of education both have a positive association with economic growth. Since education gives benefits to individuals, it is likely to see such benefits on groups of individuals as well (Stevens and Weale, 2007).

Further, education has been linked to income inequality. In very early studies, Kuznets (1956) and Adelman (1961) stated that in low educated and low-income societies, there is a more equal distribution of income – income is concentrated at this low level, which dominates the income distribution. Moving to societies with higher levels of education, the income distribution becomes more unequal because there are higher income gaps between rural and urban areas. Thirdly, when the education level is even higher and the education distribution is more equal, the distribution of income is equal again. However, many country examples give reason as to why this inverted-U theory is not realistic. In the United States of America, for example, income distribution became more equal in the 1920s-1940s, and stayed at that level until the early 1970s even though the distribution of education rapidly equalized. Despite continuous equalisation of the education distribution, after the mid-1970s, the income distribution started becoming unequal (Carnoy, 1994). Korea is another such example that contradicts the inverted-U theory.

Another macroeconomic issue where education is seen to have a role to play is in the reduction of poverty. Poverty and education are inversely related, as an educated population would earn higher incomes causing a fall in the number of individuals trapped by poverty. Education can have an effect on poverty rates of an economy in indirect ways. Parents with lower levels of education or from poorer backgrounds may give less importance to education, or are unable to assist their children with schoolwork. Secondly, if people in poor and rural areas receive lower returns to education, it could discourage them from education attainment (Brown and Park, 2002). Finally, Connell (1994) says that weak support and de-motivation could lead to less education, which in the long term causes labour market failures and high poverty rates that carry on to the next generation.

2 Henceforth, this paper refers to the Mincerian returns to education, unless stated otherwise

Poverty persistence has been seen in many countries, despite the implementation of sound education policies (Moav, 2004).

2.2 Microeconomic estimates of the returns to education

Having considered the macroeconomic evidence, it is apparent that sound education policies play a vital role in the development of a country. A considerable amount of research has been conducted on the rates of return to schooling using methods such as Ordinary Least Squares (OLS) and Instrumental Variables (IV). We will now discuss certain papers that have employed OLS estimation before moving on to the discussion of papers that have looked at the preferred method of estimation – IV.

Kingdon et al. (2008) look at several developing countries including African and Asian countries. Their estimates show that the return to primary education is, on average, 7.4 per cent, the return of lower secondary education 9.6 per cent, 12.3 per cent for higher secondary education and 19.8 per cent for tertiary education. Further results that look at other countries showed a similar trend (see Schultz, 2004) – with primary education, on average, yielding the lowest returns and tertiary education yielding the highest returns. Studies that look at the returns to education varying across sectors of employment have found that highly-educated workers in the private sector receive the highest returns, whereas the government sector rewards workers with lower levels of education (for example, this finding was seen in the Chinese economy by Li, 2003).

Evidence indicates that more educated individuals are able to earn higher wages compared to their less-educated counterparts. Jones (2001) and Temple (2001), for example, suggested that varying levels of educational attainment could explain a large part of the gaps in productivity across individuals with different levels of education. Arrow (1973) discusses that observed differences in earnings across individuals with varying levels of education may reflect differences in ability. Ability has been defined broadly as unobservable factors such as family background that affect earnings through its effect on education. Not accounting for such characteristics could lead to biased OLS estimates. The IV estimation is a way of reducing the bias from omitted variables that affect the causal relationship between education and earnings.

In order to get an unbiased estimate of the rate of return to education, we need instruments that affect an individual's education, but not their earnings. Angrist and Krueger (1991) use quarter of birth as an instrument (interacted with the year of birth in certain specifications) for education. Their results show that men in the U.S. born in the earlier quarters of the year over the period 1930 to 1959 had slightly less schooling compared to those born later in the year. This is because children born earlier in the year reach the school-leaving age of 16 at a lower grade compared to children born later on during the year. Children can leave school with less education once they reach the minimum school-leaving age. The results by Angrist and Krueger (1991) suggest that the IV estimates (which make use of the quarter of birth variables as instruments for

education) are higher than the OLS estimates of the returns to education, and in some cases, very similar to each other[3]. Bound, Jaeger and Baker's (1995) critique of the paper argues that several interactions used were weak instruments, leading to an asymptotical bias towards the OLS estimate.

Staiger and Stock (1997) used the same sample as Angrist and Krueger with the same interacted instruments to obtain IV and limited information maximum likelihood (LIML) estimates. The LIML estimates were 50 per cent higher than the corresponding OLS estimates, while the IV estimates remained higher than the OLS estimates. Bound and Jaeger (1996) argue the quarter of birth could also be correlated with differences in unobserved ability and family background - that is, some of the children born in the earlier quarters come from poorer family backgrounds, which might lead to less education and lower earnings. Card (2001) ran similar regressions using parental education as instruments to account for family background and found no proof that children born early during the year are from relatively poor households. A study by Card (1995) interacting college proximity with family background as an instrument for schooling suggested that college proximity has a bigger impact on children from poorer backgrounds.

Several other studies have used family background variables such as parents' or spouse's education as instruments (Blackburn and Neumark, 1993, 1995; Parker and Van Praag, 2006). The advantage of using such variables is that they are usually strongly correlated with education (and are available in many datasets), thus avoiding the problem of weak instruments (Bound, Jaeger, & Baker, 1995). However, the use of family background instruments has been criticised (see Psacharopoulos and Patrinos, 2004 and Trostel, et al., 2002) on the grounds that these instruments, even after controlling for education, may affect an individual's income – for example, through network effects or nepotism. Trostel et al. (2002) used spouse's level of education as an instrument after testing for the validity of the exclusion restriction required for IV – that is, spouse's education has no effect on individual earnings. Pencavel (1998) suggests that the assortative nature of marriage, where couples are more likely to share common preferences, similar behavioural traits and a similar level of education with each other, makes spouse's education a reasonable instrument for individual education.

Harmon and Walker (1995) use a change in the minimum school-leaving age in the UK that had occurred in 1947, which increased the school-leaving age from 14 to 15 years of age. Their results suggested that there is a significant and downward bias in the OLS estimate of the years of schooling, in relation to earnings (returns to schooling). Such an instrument is useful when analysing the return to economically disadvantaged groups, because they are less likely to attain education beyond the compulsory number of years in the absence of such a policy. And therefore, such groups may be more influenced by the policy change compared to higher income groups.

3 The academic year in the U.S. runs from January to December. In Sri Lanka, government schools follow the same system whereas private/international schools commence the academic year in September.

Card (2001) reviews several studies that use IV to estimate the return to schooling. In most studies, the IV estimate of the return to schooling is higher than the corresponding OLS estimate, usually by 20 per cent or more. One explanation given for this is measurement error (observed level of schooling differs from the true level attained) that causes a downward bias in OLS estimates (Griliches, 1997; Angrist and Krueger, 1991). Card (2001) suggests that this explains roughly 10 per cent of the gap between the IV and OLS estimates, arguing it is unlikely that so many studies found such large gaps primarily due to measurement error. A second explanation by Card (2001) is that the upward bias could be due to unobserved differences in characteristics between the treatment and comparison groups used in IV – as grouping may emphasize the upward bias even further by reducing the variance in the education variable by more than it reduces the covariance of the education variable with the bias terms (see Card for a detailed explanation). This is especially problematic in a quasi-experimental approach.

A third explanation, as suggested by Ashenfelter et al. (1999) is "reporting/publication bias". This arises because there is a tendency for studies to report results that are statistically significant and have an effect thus leading to education estimates that are correlated with sampling errors. If in turn, these are correlated with other variables, findings about the returns to education could be biased. After adjusting the estimates for reporting bias, Ashenfelter et al. (1999) found some of the returns to be much lower than what was actually reported. The final explanation, given in Card (2001), is the possibility of underlying heterogeneity in the returns to education. This would mean that the IV estimates based on supply-side features such as the minimum school-leaving age or the geographic proximity of schools have a tendency to improve returns for a group of individuals who have relatively high returns to education. These supply-side features affect schooling choices of people who would otherwise have relatively less years of schooling. Thus, less-educated individuals affected by supply-side innovations have relatively high marginal returns to education reflecting their higher-than-average costs of education, rather than because of low ability that would restrict the returns to schooling. In such an instance, the "local average treatment effect" (LATE) implies that supply-side features could produce IV estimates of the returns to schooling above the corresponding OLS estimates and the average marginal return to schooling.

Having discussed the IV approach, the various instruments used in the literature and the possible explanations for differing OLS and IV estimates of the returns to education, we will now discuss some of the papers that have looked at the rates of returns to education in Sri Lanka.

2.3 Returns to education in Sri Lanka

Despite the availability of data, to our knowledge[4], limited research has been done in this field for Sri Lanka. Aturupane (2011) discussed the link between the supply of and returns to education, at different levels[5]. The paper observed a high supply of "basic educated human capital" in Sri Lanka and relatively low returns to education at the primary level (approximately 2 per cent). The pattern is reversed at the secondary and tertiary levels where the supply of labour with higher education is low but higher returns are observed (13 to 21 per cent). In order to increase the supply of more educated labour, Aturupane (2011) argues that there is a need for greater emphasis on the development of secondary schools, and schools located in rural and estate areas which have less-qualified staff, poorer facilities and managerial systems.

With universal education, public expenditure on education per student is lower than other middle income countries in Latin America and East and South Asia (Aturupane, 2011) – Public expenditure per student in 2011 was 8.5% per capita GDP compared to expenditure per student in India of 11.8% per capita GDP (World Bank, 2014). Figures 2 (a) and (b) both show where Sri Lanka stands amongst other countries in terms of public expenditure invested in education and the expenditure per student, respectively. The countries used in comparison are middle-income countries and other comparable countries that have characteristics similar to Sri Lanka (as given by Aturupane, 2011).

Figure 2 (a) represents the total public spending on education as a percentage of the Gross Domestic Product for 2010. Sri Lanka spends the least amount of public expenditure on education in comparison to many other countries[6]. Figure 2 (b) looks at education expenditure per student as a share of GDP per capita[7] - this figure for Sri Lanka also seems to be comparatively lower than several other countries, suggesting that Sri Lanka under-invests in education relative to other countries. Aturupane (2011) stated that Sri Lanka's low public investment in education is associated with weaker performance in the Trends in International Mathematics and Science Study (TIMSS); whereas countries with high public spending in education performed well in the TIMSS.

Since 1938, Sri Lanka has had a free education policy from the level of kindergarten up to undergraduate studies. Two reasons in favour of the provision of such subsidised education have been related to equity and efficiency. Education enriches an individual's capacity to earn more income leading to greater equity in an economy. If not subsidised, education will be restricted to the rich who can afford it, leading to a continuous cycle of "inter-generational poverty transmission" (Ranasinghe and Hartog, 1997).

4 Based on a search in EconLit

5 The paper used World Bank estimates on the returns to education – no further information is given on the year that the analysis was conducted, etc.

6 Similar patterns were seen for the previous years – 2008, 2009 and 2011

7 Data was either from the year 2010 or a year close to it when the data was not available for all the countries

***Figure 2(a)**: Public spending on education (as a % of GDP) in the year 2010*

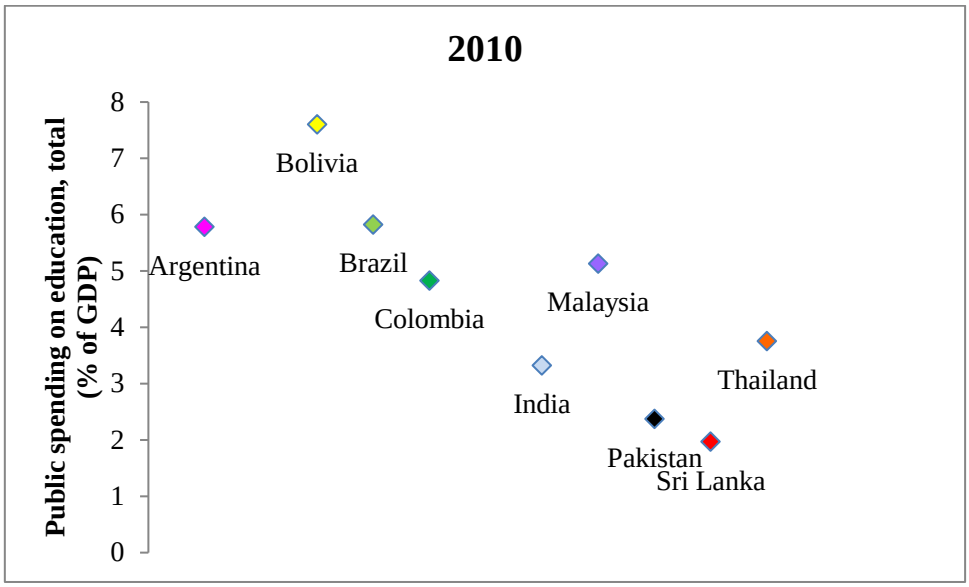
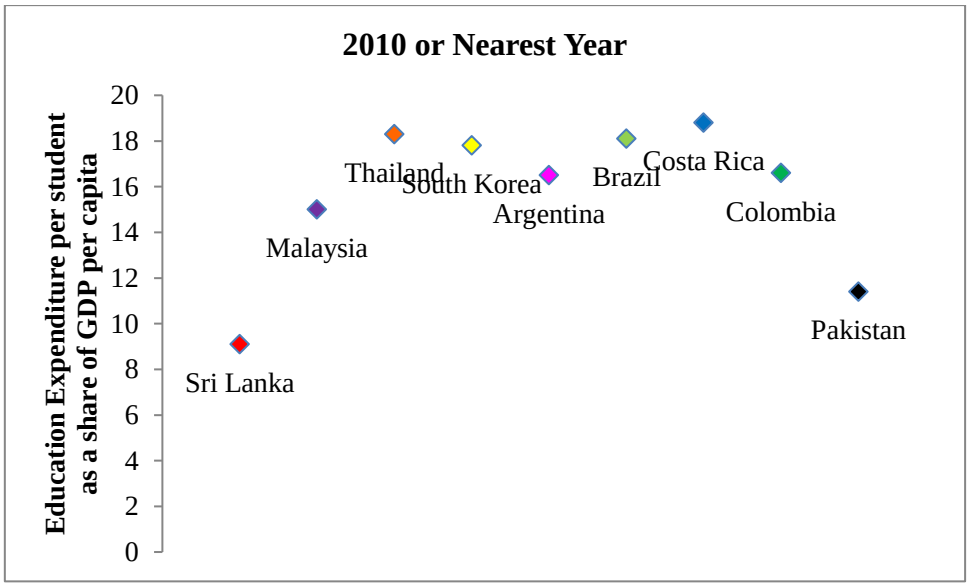


***Figure 2(b)**: Education Expenditure per student (as a share of GDP per capita)*



The second argument in favour of free education is efficiency. It is believed that education brings about positive externalities – the acquired skills can benefit the wider society via higher productivity and lower unemployment rates, for example. However, it is argued people may make individually optimal decisions which are not socially optimal. In comparison to other developing countries, Sri Lanka's free education system up to, and including, the undergraduate level is unique. The country, having transformed from a low-income to a middle-income country, is ahead of many of its neighbours with regard to adult literacy and enrolment rates (World Bank, 2005). Recent statistics have shown that school survival rates till grade 9 have risen from 78 per cent in 2005 to 91 per cent in

2011. Primary school enrolment rates have been 99 per cent since 2010 (World Bank, 2014).

However, Ranasinghe and Hartog (2002) find that children from well-off families are likely to stay longer in school and have greater potential to pass the final exams, compared to their poorer counterparts. The paper suggests that free education has been unsuccessful in erasing the "family background effect" on school attendance; children from richer families benefit more from the system of free education than poor children. They argue free education has worsened the education and income distribution by subsidising the education of wealthy households.

Having looked at several South Asian countries, Riboud et al. (2007) found that progress over the years was uneven. In terms of the share of the population that has completed at least primary education, Sri Lanka was an outlier compared to other South Asian countries such as India, Bangladesh, Pakistan, Bangladesh and Nepal; more than 70 per cent of people born in the late 1940s had completed at least their primary education, and this trend continued for the next 40 years, after which universal primary education was introduced. However other countries that had the lowest levels of education started catching up to the front-runners with regard to universal primary education.

Looking at lower secondary education, Sri Lanka has been ahead of many of its neighbours in terms of the share of the population who have completed this level of education. Moving on to upper secondary education, Sri Lanka is no longer an outlier. The country's focus on basic education has left it much less focused on the higher levels of schooling. At all levels of education, the country has a larger proportion of girls than boys completing each level of education (up until higher secondary). Riboud et al. (2007) also look at the wage gains from completing each level of education using a Mincerian wage model, and find a significantly greater wage gain from completing secondary and tertiary education, compared to the gain from completing primary education.

Workers located in cities and urban areas command higher wages than those in rural or estate areas. Poverty in Sri Lanka is believed to be a "largely rural phenomenon" (Datt and Gunerwardena, 1997), with nearly half of the poor reliant on agriculture for their living, while another 30 per cent rely on other non-agricultural rural jobs. The estate sector accounts for the second largest share of national poverty, while the urban sector accounts for the least. Recent estimates between the years 2006 and 2009 show that poverty incidences are greatest in the estate sector, followed by the rural and urban sectors.

Gender differences in wages have been observed over the years in many countries. In Sri Lanka, wage differentials of approximately 25 per cent have been noted, despite women attaining higher levels of education in comparison to men (Gunerwardena, 2002). This is lower in comparison to other South Asian countries such as India and Pakistan where large wage gaps of 50 per cent and even 300 per cent respectively have been noted

(Riboud et al., 2007). Explanations for such differences include the type of occupation, sector or the degree of discrimination. Riboud et al. (2007) showed that the returns to higher secondary and tertiary education in particular have increased, and comparatively more for women than for men. Women appear to receive lower primary returns than men in all three years of the analysis (1992/93, 1997/98 and 2001/02). At the secondary and tertiary levels of education however, the results have changed, with women receiving higher returns to education than men.

Women may receive higher returns, but Gunerwardena's study concludes that the wage gap remains in favour of men. If women are disproportionately less educated, it will have an impact on poverty – literature shows that female-headed households, the illiterate and less educated are more likely to be trapped by poverty (Connell, 1994; Awan et al, 2011). Recent studies suggest that although women may have restricted access to higher levels of education and low labour force participation in certain countries, once they overcome the obstacles, they do well in the labour market (Riboud et al., 2007). Women with higher levels of education would choose to participate in the labour market.

Educational attainment and the returns to education vary by level of education and across different sub-groups. The papers discussed above in the context of Sri Lanka have employed OLS to estimate the rates of returns to schooling. In order to compare OLS estimates to IV, this chapter will examine both approaches to analyse the returns to schooling in Sri Lanka in 2009/10, which was shortly after the end of a civil war. This will allow us to identify whether the OLS estimates tend to be upward or downward biased. The next section will explore the data used for the analysis.

3. Data and methods

3.1 Data

The data used in this study comes from the Sri Lankan Household Income and Expenditure Survey (HIES) 2009/2010. This survey comprises of a sample of 17,182 households (73,396 individuals), excluding the Northern and Eastern provinces. The excluded provinces consist of 3,511 individuals in either wage or self-employment. Due to partial data collection in these provinces that commenced at different points throughout the year, they have been excluded from the analysis[8]. The HIES includes variables on demographic, health and education factors, and a detailed breakdown of individual income and expenditure. Obtaining information on income is a difficult task (people are reluctant to disclose such information and usually under-report). This survey attempted to address this issue by gathering data at the individual level by the six income categories. The reasoning was that this was a less intrusive question and would illicit more accurate data.

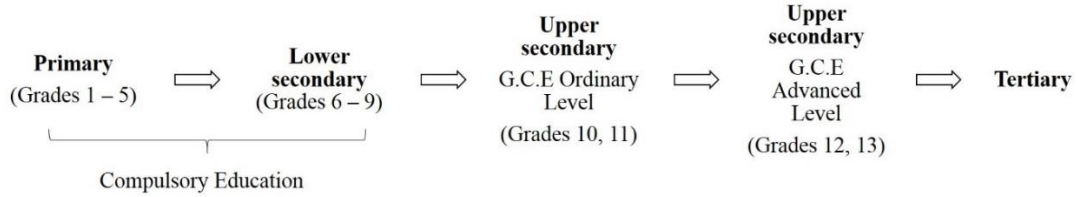
8 The returns to education are lower when all districts are included – for the full sample, as well as for the sub-samples by gender and type of employment.

The following are the six sub-divisions of income:

(a) paid employment;
(b) agricultural activities (including paddy and other seasonal crops);
(c) other agricultural activities (including tea, rubber, fish, eggs, fruits, etc.);
(d) non-agricultural activities (such as mining, construction, trade, transport, etc.);
(e) other income (pension payments, disability/relief payments, dividends/interest);
(f) windfall income (bank loans and repayment of loans given, lottery, sales of assets, etc.)

For the purposes of this study, non-earned income including other income (e) and windfall income (f) are excluded from the analysis. Therefore, earned income is used as the key variable in the analysis – this includes income from paid employment, agricultural, other agricultural and non-agricultural activities. The education system in Sri Lanka (as shown in Figure 3) is divided into primary, secondary and tertiary education. Grades 1-5 are considered primary education, with school starting at the age of 5; lower secondary covers grades 6-9; upper secondary covers grades 10-13, with university education being labelled tertiary[9].

Figure 3: Educational system in Sri Lanka



Tables 1 and 2 summarize key descriptives split by (i) gender, and (ii) sector/type of employment (agricultural and non-agricultural private and government workers). The agricultural and non-agricultural private workers include self-employed individuals and wage earners. Looking at the characteristics of men and women in the labour force in Table 1, men earn more than women on average. Men and women appear to have similar years of schooling, although women in the labour force are less likely to be married, and more likely to be working in professional occupations.

Table 2 suggests that the average earnings are lowest for agricultural workers and highest for government workers. A possible explanation for private sector workers reporting lower earnings relative to government sector employees could be due to formal and informal private sector workers being aggregated in the analysis. Unfortunately, it is not possible to disaggregate the two groups due to data constraints in the HIE survey series.

9 There is no information available in the survey on grade repetition which may cause an over-estimate. The World Bank shows that a small percentage (1 per cent) show grade repetition at the primary level, hence it may not lead to biased estimates.

Further, the summary statistics suggest that agricultural workers are older, on average. Agricultural workers reported the lowest levels of education while non-agricultural government sector workers are more educated, on average, with 11.27 years of schooling. 40 per cent of the government sector comprises of women, whereas only 26 and 29 per cent are women in the agricultural and private sectors, respectively. 94 per cent of agricultural workers reside in rural areas, while 30 per cent of non-agricultural workers reside in such areas. The next section of the chapter will examine the methods used to analyse the rates of returns to education for these different sub-groups.

Table 1: Summary statistics for the pooled sample, men and women

Variables	Full sample	Men	Women
Earnings per month (in rupees)	13,833	14,805	11,781
Individual characteristics:			
Age	42	42	41
Years of schooling	8.78	8.75	8.86
Sinhalese	0.78	0.78	0.77
Female	0.30		
Married	0.75	0.80	0.63
Urban	0.24	0.24	0.22
Occupations:			
Managers	0.07	0.07	0.05
Professionals	0.07	0.05	0.14
Technicians	0.10	0.11	0.08
Clerks	0.04	0.03	0.06
Services	0.07	0.08	0.04
Craft	0.14	0.14	0.15
Operators	0.08	0.10	0.03
Elementary	0.23	0.24	0.21
Armed forces		0.01	
Agriculture and other	0.20	0.19	0.24
Industries:			
Primary	0.30	0.29	0.32
Secondary	0.22	0.22	0.24
Tertiary	0.48	0.49	0.44
Number of observations	*23,414*	*16,282*	*7,132*

Table 2: Summary statistics for agricultural and non-agricultural private and government workers

Variables	Agriculture	Private non-agriculture	Government non-agriculture
Earnings per month (in rupees)	7,800	12,896	22,568
Individual characteristics:			
Age	49	40	41
Years of schooling	7.66	8.48	11.27
Sinhalese	0.93	0.73	0.81
Female	0.26	0.29	0.40
Married	0.80	0.71	0.82
Urban	0.06	0.28	0.27
Number of observations	*4,420*	*15,123*	*3,871*

3.2 Ordinary Least Squares

This paper estimates the earnings function at the individual level, and the production function at the household level. To estimate the *earnings function*, the sample is decomposed in two ways; gender and sector/type of employment – agricultural and non-agricultural activities in the private and government sectors. Private sector employees and employers together comprise of private non-agriculture. Similarly, government and semi-government[10] workers together comprise of government non-agriculture. We focus on the individual level rather than simply focussing on the characteristics of the head of the household[11]. Therefore, the unemployed and those who are not employed (students, those who are retired or unable/too old to work) were eliminated from the analysis.

The dependent variable is the natural logarithm of earned income, calculated at the monthly level. This variable is obtained in the following manner. For wage earners, income includes basic pay, bonuses/arrears, tips, commissions and overtime pay. For the self-employed, income is calculated as output less input and private consumption. The HIES obtains these figures on a monthly basis.[12] Some individuals (2012 observations)

10 Semi-government organizations are government-owned; however they may be run by the private sector and have financial objectives, apart from policy objectives.

11 This will account for children in a family (younger generation) who may be more educated than their parents and contribute to the household.

12 There is no data available on hours worked in the HIE surveys. Looking at monthly earnings in chapter 4, instead of hourly earnings, we observed that the returns to education are lower with the former as the dependent variable (with the exception of self-employment where the returns to education are higher when monthly income from self-employment is used instead of daily income). Therefore, we expect our findings in this chapter on the returns to education to be higher if hourly earnings were used as the dependent variable in the case of wage earners; On the other hand, we expect the returns to education to be lower if hourly earnings are used for the self-employed.

receive more than one source of income. To account for this, we use the aggregate income from both activities. Individuals are grouped into agriculture or non-agriculture (private or government) categories based on their primary occupation.

The following Mincer earnings function is estimated:

$$log(Y_i) = \beta_0 + \beta_1 S_i + \beta_2 W_i + v_i \quad (1)$$

where *log* (Y_i) is the natural log of monthly earnings. The term "S" in Equation 1 represents a vector of the education variables – no education, primary, lower secondary, upper secondary and tertiary education. Tertiary education[13] and not having obtained any form of education are measured as binary variables, giving a value of one if the individual has obtained any tertiary education (or no education), or zero otherwise. The other education variables (primary, lower secondary and upper secondary education) capture the years of each level of education obtained by an individual.

"W" is a vector that captures the individual characteristics. The gender is included through the variable called "Female" which has a value of one if the individual is female or zero if male. The "Urban" dummy variable gives a value of one if the individual is from an urban area (this variable will be changed later on in the production function to a "Rural" dummy variable[14]). The "Married" dummy variable gives a value of one if an individual is married, else zero. Ethnicity is captured by the dummy variable "Sinhalese" which gives a value of one if an individual is Sinhalese, or zero otherwise; over 50 per cent of individuals in the sample are of this ethnic background. Age, age^2 and age^3 represent the age of the individual, age-squared and age-cubed, respectively. The reason for including the square and cubic terms is because it is believed that age and earnings share a non-linear relationship. Finally, the last term in Equation 1 (v_i) represents the stochastic error term.

A further estimation of the earnings function is carried out in order to get a better understanding of how the returns to employment vary across different types of occupations and industries We now have additional regressors measured as dummy variables: "X" represents a vector of the industry classifications (primary industries such as forestry, fishing, mining and quarrying, secondary industries such as manufacturing, electricity, water and gas supply, construction, and tertiary industries such as trade, transportation, finance and insurance, real estate, education, administration, public administration and defence, professional, scientific and technical activities, health and social work, arts, entertainment and recreation, work for the household, and activities of

13 Tertiary education was taken as a binary variable since the exact years in tertiary education is not fixed and therefore, unknown. However, it will be assumed that an individual gets three years of tertiary education for the purposes of instrumental variables

14 In the production functions, we focus on the self-employed; and self-employed agriculture is pre-dominantly in the rural sector. We could use the urban sector as the default category instead, in which case the coefficient will change signs.

extra-territorial bodies); "Z" represents a vector of the occupation classifications (high-skilled white collar jobs such as managers, professionals and technicians, low-skilled white collar jobs such as clerks and service workers, high-skilled blue collar jobs such as craft and other trade workers, low-skilled blue collar jobs such as machine operators, assemblers and elementary workers).

In the next part of the OLS estimation, we estimate a *production function* at the household level for the self-employed in agricultural and non-agricultural activities. Non-agriculture in this scenario includes activities such as mining, construction and trade, transport, manufacturing and services – it is not an aggregation of the self-employed in the private and government sectors from the earnings function discussed above. The purpose of carrying out the production function at the household level is because the sample contains unpaid family workers (201 observations) who are dropped out of the previous regression, as they do not report earnings. This category includes the head of the household, or any other family member such as the spouse, child, or any other relative.

Further, unlike with wage income it is less clear as to how income/output from self-employment is assigned; if income/output is from a family enterprise, assigning it to one individual is problematic as it may be held collectively. Therefore, the production function is set up so that the output produced is from the "enterprise" – thus, enterprise characteristics are used. In this instance, the variable "primary education" for example, represents the average years of primary education obtained by all employed individuals residing in the household[15]. Lower and upper secondary education has been measured in a similar manner. In the production function, the main variable of interest is the natural log of output, rather than income.

We employ the following translog production function (Jacoby, 1992; Appleton and Balihuta, 1996), to estimate the effects of the factors of production, X_j $(j = 1,2,3)$, on output "Q" for farm "h":

$$(2) \qquad log\_Q_h = \alpha_0 + \sum_{j=1}^{3} \alpha_j logX_{jh} + \sum_{j=1}^{3}\sum_{k=1}^{3} \delta_{jk}\, logX_{jh} logX_{kh} + +\gamma_1 S + \gamma_2 W + v_i$$

The reason for this separation of the agricultural from the non-agricultural households is because it is possible that the returns to education in agriculture may be lower. Further, the technology being used differs between these two groups; for example, land is essential for agriculture, but no information on land use has been recorded in the survey for non-agriculture. The production functions are generated in the following way. Inputs include raw materials and capital (included together as "inputs"), land[16] and labour.

15 Due to data limitations, we are unable to identify whether these individuals work in the enterprise or elsewhere.

16 Land is taken in acres, roods and perches. For convenience, all these values were converted to acres and summed up: 1 rood = 0.25 acres; 1 perch = 0.00625 acres

Unfortunately, there is no detailed breakdown of the cost of inputs, but the sum of capital and raw materials (such as seeds and fertilizer)[17].

The inputs are calculated by substituting $(logX_{jh} - logX_j^*)$ for $logX_{jh}$, where X_j^* represents the sample mean for each factor of production. This substitution is done since the individual coefficients are difficult to interpret in a translog function because of second order terms. By making this change, it is now implied that the coefficients on the scaled "X" terms $(logX_{jh} - logX_j^*)$ are the elasticities of production evaluated at the sample means. The next term in Equation 2 represents the interactions between the factors of production. The "S" term represent a vector of the schooling variables and "W" represents a vector of other enterprise-level characteristics – average age, age-squared and age-cubed of all working individuals and the proportion of employed individuals who are women/Sinhalese/married. The stochastic error term is the final variable given in Equation 2.

3.3 Instrumental Variables

As discussed in Section 2.2, a potential issue with OLS is that "S", representing the aggregate years of schooling[18], could be endogenous. If so, "S" may be correlated with the error term ε_i which represents other omitted factors that determine earnings. This would cause the OLS estimate to be inconsistent. Hence, the use of an instrumental variable (IV) approach would help solve many related issues such as measurement error and omitted variable bias. By the use of instruments, it can isolate the part of "S" that is uncorrelated with ε_i

The IV regression is as follows[19]:

$$log\_Y_i = \beta_0 + \beta_1 \widehat{S}_i + \beta_2 W_i + \varepsilon_i$$

where $\hat{S}$ is a predicted value of schooling that has been obtained from a first-stage regression of "S" on a constant term, the instrument(s) (represented by "Z") and the exogenous variables "W":

$$\hat{S}_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 W_i + e_i$$

The instruments (Z) that will be used are the spouse's education (in years) and a change in compulsory schooling laws. These instruments should satisfy the following conditions:

---

17 Given that there is no data available on the capital stock in isolation, it is difficult to distinguish the spending on capital from the spending on raw materials

18 Levels of education (primary, secondary and tertiary) could not be used in IV since an over-identified model was required; and a sufficient number of instruments were not available.

19 The IV regression is run using the STATA command "ivregress" with the two-stage least squares estimator.

- Instrument exogeneity – The instruments should not have any partial effect on wages after controlling for the years of schooling and other omitted variables. The instruments should be uncorrelated with any other determinants of the dependent variable. This is the exclusion restriction.

$$Cov(Z, \varepsilon_i) = 0$$

- Instrument relevance – The instruments must be related to the years of schooling, which is the main explanatory variable. That is, a first stage should exist.

$$Cov(S, Z) \neq 0$$

Two instruments are used in this study – (1) the implementation of compulsory schooling and (2) spouse's education. Following Harmon and Walker (1995), we make use of the introduction of compulsory schooling in the case of Sri Lanka. The change in compulsory schooling laws occurred in the year 1997 when it was made compulsory for children between the ages of 5 and 14 to attend school. Before 1997, it was not compulsory. While it is compulsory to complete 9 years of schooling after 1997, out of the 3,383 observations affected by the policy, 78 per cent completed compulsory education. Almost all the individuals who did not complete the years of compulsory education are from rural areas. A possible explanation could be the lack of access to education facilities for these individuals. Out of the 20,031 individuals who were not faced with the policy, 50 per cent completed 9 years of education – this suggests that the education distribution has been affected by the implementation of compulsory schooling laws.

In this paper, the change is measured through the variable "policy". It will have a value of one if a child is born after 1983 (they will be 14 years or younger in 1997), or zero otherwise. Since this survey was carried out in 2009/2010, the policy change denotes a value of one to those individuals who are 26 years or younger during the time of the survey. The second instrument considered to have an impact on an individual's level of education is spouse's education. It is believed that individuals get married to people with similar behavioural traits, preferences and levels of education (Pencavel, 1998)[20]. We follow other studies such as Blackburn and Neumark (1993, 1995), Parker and Van Praag (2006) and Trostel et al. (2002) that have used family background (parents/spouse) as instruments.

20 While some studies (Psacharopoulos and Patrinos, 2004; Trostel, et al., 2002) raise concerns as to whether family background is possibly related to earnings, we choose to proceed with the instruments if the exclusion restrictions are not violated.

3.3.1 Tests for weak instruments, over-identification and endogeneity of education

In estimating IV it is important to ensure that none of the instruments are weak, the instruments are valid, and finally that the regressor (years of schooling) is indeed endogenous.

The first is the test for weak instruments – if the change in compulsory schooling laws and spouse's education (that is, the instruments) are not strongly correlated with the years of schooling (the endogenous regressor), they are weak instruments. This would cause a bias in the IV estimators. Since there is a single endogenous regressor, weak instruments were tested for by looking at the joint significance of the instruments in the first stage. Staiger and Stock (1997), among others, suggest that an F-statistic greater than 10 implies that the instruments are not weak; the null hypothesis for weak instruments is rejected.

The essential requirement is that there should be at least as many instruments as endogenous explanatory variables. In such a case, the model is just identified. Instead in this case, there are more instruments (2) than endogenous variables (1); hence the model is over identified. It is important to test for the over-identifying restrictions (the number of surplus instruments, that is, 1), which is the second test to ensure that the instruments are valid. The test for over-identification suggests that if the null hypothesis of instruments' exogeneity cannot be rejected, there are no offending instruments.

The final test will check whether the regressor (years of schooling) is endogenous or exogenous. For this, the robust version of the Hausman test is used. Firstly, regressing the years of schooling on the constant term, the exogenous regressors (W), and the excluded instruments (Z), a residual is obtained. Then, a Mincer wage function using OLS with robust standard errors is estimated, adding the residual obtained from the previous step as an additional regressor. Finally the null hypothesis that the coefficient of the residual is equal to zero (implying that the regressor is exogenous) is tested. The null is rejected if the "years of education" is an endogenous regressor, and the use of IV is justified. If the null is not rejected, the IV estimates are still consistent, but they are inefficient relative to OLS which may be preferred instead.

In all specifications, the instruments pass the tests for weak instruments and over-identification – with F-statistics greater than 10, and p-values greater than 5 per cent respectively. Hence the instruments are valid. The Hausman test suggests that the null hypothesis for exogeneity of education is rejected. Therefore, education must be treated as an endogenous regressor, and treating it as exogenous will lead to biased OLS estimates[21].

[21] Following Angrist and Krueger (1991), the quarter of birth was also used as an instrument. Since the Sri Lankan education system is similar to the American one (where the academic year starts in January), quarter of birth seemed the most appropriate as this is used by Angrist and Krueger (1991) and allows for comparison between the two studies. However, the instrument did not pass the test for weak instruments. A possible explanation for this could be because the academic year runs differently across public and private schools; the academic year commences in January in public schools,

4. Results and discussion

4.1 IV (in comparison to OLS)

This section compares the results from two approaches – IV and OLS. The aggregate years of education is used as the main variable of interest, with the spouse's level of education and a change in compulsory schooling laws that occurred in 1997 (a policy change that gives a value of one to those born after 1983) as the instruments. The first-stage IV regression results are given in Appendix 1. The earnings function is estimated separately for men and women in Section 4.1.1 and for agricultural and non-agricultural private and government workers in Section 4.1.2.

4.1.1 OLS versus IV: men and women

Firstly we estimate the earnings functions for men and women in the labour force. Tables 3 and 4 present the results for the OLS and IV estimations, respectively. The tables present the Mincerian specification (1) along with the full specification (2) includes accounting for the occupation and industry of the individual.

For the purpose of comparing IV with OLS, we focus on the Mincerian returns (given by specification (1) in the tables). For men, the IV estimate for the return to education is 12.6 per cent, higher than the OLS estimate of 8.6 per cent. Similarly for women, the IV estimate is 14.8 per cent while the OLS estimate is 8.9 per cent. This suggests that the OLS estimates are downward-biased. Card (2001) reviewed several studies where this downward bias of the OLS estimate was present, as discussed in the literature review. There could be several reasons for this downward bias. Measurement error, not being able to measure unobservable characteristics such as ability and family background, underlying heterogeneity in the rates of returns to education, and even reporting bias are some of the possible explanations. Other studies that found higher IV estimates compared to the OLS estimates include the papers by Angrist and Krueger (1991), Staiger and Stock (1997), Harmon and Walker (1995). Comparing the (IV) returns for men and women, women receive a higher return of 14.8 per cent for every additional year of schooling compared to 12.1 per cent for men.

whereas it commences in September in private schools. (see results using quarter of birth and parents' education in Appendix 2).

Table 3: OLS estimates for men and women

Dependent variable: log(earnings)	Men		Women	
Independent variables	(1)	(2)	(1)	(2)
Years in education	0.086***	0.052***	0.089***	0.036***
	(37.60)	(19.39)	(28.81)	(8.94)
Age	0.131***	0.115***	0.091***	0.076***
	(9.70)	(8.66)	(4.80)	(4.01)
Age^2	-0.002***	-0.002***	-0.002***	-0.001***
	(-7.28)	(-6.51)	(-3.59)	(-3.04)
Age^3	$1.04x10^{-5}$***	$9.23x10^{-6}$***	$7.55x10^{-6}$**	$6.47x10^{-6}$*
	(4.51)	(4.06)	(2.26)	(1.95)
Urban	0.226***	0.119***	0.245***	0.067***
	(14.97)	(8.22)	(9.06)	(2.60)
Married	0.223***	0.215***	-0.029	0.001
	(10.89)	(10.79)	(-1.11)	(0.04)
Sinhalese	0.082***	0.124***	-0.158***	-0.151***
	(5.04)	(7.89)	(-5.87)	(-5.45)
Occupations				
Managers		0.872***		0.705***
		(19.13)		(9.12)
Professionals		0.801***		1.020***
		(17.09)		(16.66)
Technicians		0.591***		0.735***
		(15.17)		(12.19)
Clerks		0.554***		0.918***
		(12.69)		(15.45)
Services		0.386***		0.463***
		(10.01)		(7.13)
Craft		0.345***		0.019
		(9.21)		(0.33)
Operators		0.382***		0.453***
		(10.14)		(6.21)
Elementary		0.192***		0.235***
		(6.03)		(6.05)
Armed forces		0.967***		1.484***
		(20.88)		(11.30)
Industries				
Secondary		0.137***		0.214***
		(5.42)		(4.86)
Tertiary		0.097***		0.140***
		(4.09)		(3.42)
Constant	6.213***	6.347***	6.692***	6.920***
	(35.17)	(5.93)	(27.18)	(27.58)
N	16,282		7,132	

* indicates significance at the 10% level; ** 5% level; *** 1% level. The z-statistics are in parentheses. Default categories are: rural, not married, non-Sinhalese, agriculture, primary industry

Table 4: IV estimates for men and women

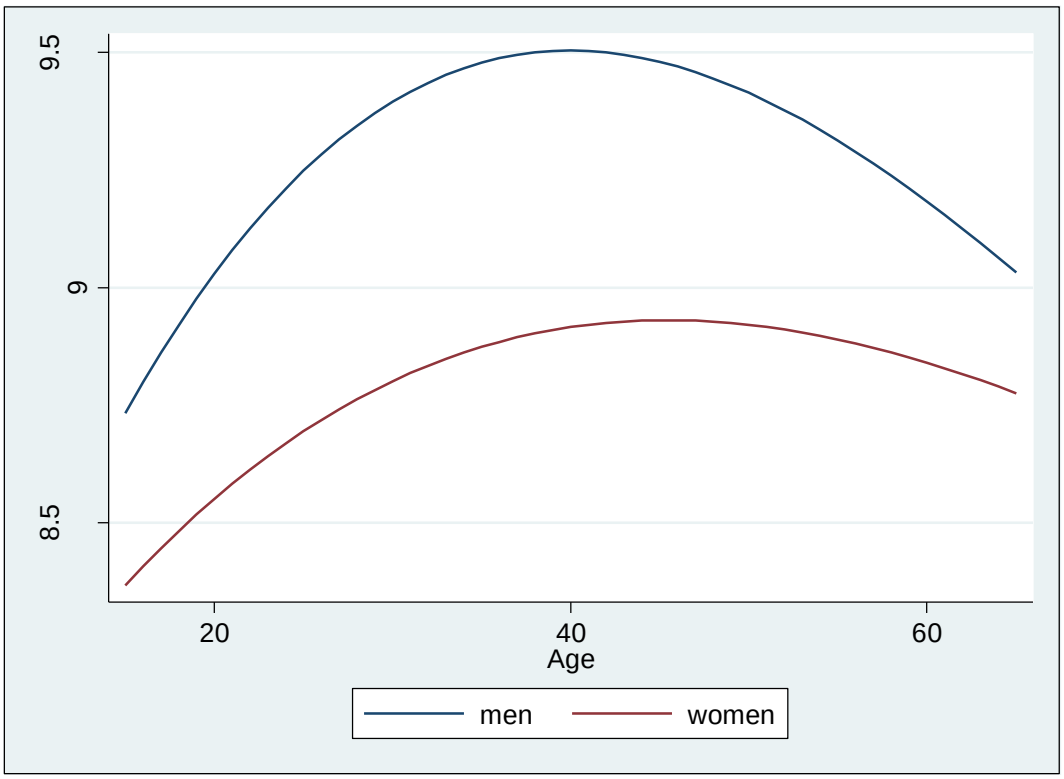
Dependent variable: log(earnings)	Men		Women	
Independent variables	(1)	(2)	(1)	(2)
Years in education	0.121***	0.088***	0.148***	0.104***
	(27.44)	(14.58)	(24.09)	(9.87)
Age	0.126***	0.115***	0.071***	0.066***
	(9.37)	(8.76)	(3.64)	(3.49)
Age^2	-0.002***	-0.002***	-0.001**	-0.001**
	(-6.92)	(-6.50)	(-2.34)	(-2.38)
Age^3	$9.75x10^{-6}$***	$9.14x10^{-6}$***	$4.34x10^{-6}$	$4.87x10^{-6}$
	(4.24)	(4.06)	(1.26)	(1.46)
Urban	0.177***	0.101***	0.145***	0.040
	(11.28)	(6.92)	(4.91)	(1.49)
Married	0.246***	0.236***	-0.003	0.022
	(11.88)	(11.71)	(-0.12)	(0.81)
Sinhalese	0.001	0.058***	-0.405***	-0.342***
	(0.06)	(3.20)	(-10.86)	(-8.52)
Occupations				
Managers		0.797***		0.572***
		(16.95)		(7.15)
Professionals		0.615***		0.663***
		(11.22)		(8.21)
Technicians		0.518***		0.555***
		(12.71)		(8.44)
Clerks		0.450***		0.676***
		(9.68)		(9.71)
Services		0.343***		0.360***
		(8.74)		(5.29)
Craft		0.350***		-0.022
		(9.29)		(-0.37)
Operators		0.360***		0.406***
		(9.46)		(5.42)
Elementary		0.241***		0.320***
		(7.38)		(7.45)
Armed forces		0.916***		1.312***
		(19.37)		(10.50)
Industries				
Secondary		0.098***		0.143***
		(3.73)		(2.98)
Tertiary		0.056**		0.080*
		(2.23)		(1.82)
Constant	6.011***	6.079***	6.541***	6.614***
	(33.75)	(33.70)	(25.85)	(25.90)
N	16,282		7,132	
Test for weak instruments	**Pass**	**Pass**	**Pass**	**Pass**
F-statistic:	2881.4>10	1589.14>10	1287.9>10	491.88 >10

Test for over-identification	**Pass**	**Pass**	**Pass**	**Pass**
p-value:	0.08>0.05	0.07>0.05	0.15>0.05	0.49>0.05
Test for exogeneity	**Reject**	**Reject**	**Reject**	**Reject**
p-value:	0.00<0.05	0.00<0.05	0.00<0.05	0.00<0.05

* indicates significance at the 10% level; ** 5% level; *** 1% level. The z-statistics are in parentheses. Default categories are: rural, not married, non-Sinhalese, agriculture, primary industry

Figure 4 shows an inverse U-shaped relationship between the age variables and (predicted) earnings[22] for both, men and women. Human capital theory explains this concave relationship as follows – at early stages of the career, one reason for the steep and positive relationship between wages and age is because of "learning on the job". As an individual gets older and accumulates less human capital, the growth of earnings slows down. On average, women get the highest earnings at the age of 45, and men get to their peak level of earnings at the age of 40. After these peak points, predicted earnings are increasing at a decreasing rate. Across the sexes, predicted earnings are higher for men at every age in comparison to women.

Figure 4: Age-Earnings profiles for men and women



22 Predicted earnings were calculated in the following manner using the IV estimates. First, we calculated the sum of the vector of all coefficients multiplied by the respective means – for all variables with the exception of age, age-squared and age-cubed (1). Next, age, age-squared and age-cubed were multiplied by their respective coefficients (2). Finally, we sum up the figures obtained for (1) and (2) with the constant term obtained in the regression – which represent the predicted earnings. Predicted earnings were then plotted against age to obtain the curves presented in Figures 4 and 5

The coefficients on the urban dummies are positive for both sexes indicating that men and women in urban areas receive higher returns than their rural counterparts. Married men receive higher returns, on average, than unmarried men. In the literature, one explanation for this is the selectivity effect (Gwartney and Stroup, 1973) – men who earn higher wages are more likely to get married. On the other hand, our analysis suggests that married women do neither better nor worse than their unmarried counterparts. The coefficient on the ethnicity variable is negative for women implying that the returns are higher for non-Sinhalese women compared to their Sinhalese counterparts; whereas there appear to be no ethnic differences across men.

The first-stage IV estimation results for men and women are presented in Appendix 1. Women from urban areas attain more education compared to their rural counterparts, whereas there are no significant urban-rural differences in education for men. Women from a Sinhalese background attain substantially higher levels of education (2.94 years) whereas Sinhalese men also have an advantage even though it is not as large in size (0.91 years). Again, the two instruments that are used to deal with the issue of endogeneity in education are the policy change and spouse's level of education. The policy instrument measures the impact of the introduction of compulsory schooling until the age of 14, which occurred in 1997. This has a relatively large and positive impact on women's education (0.92), whereas for men, the effect is not as large (0.33). Spouse's education is positively related to individual education, for both, men and women.

The IV estimates for the male-female sub-groups passed the tests for weak instruments and over-identification; implying that the instruments are not weak and are exogenous, hence they are valid. The Hausman test suggests that the years of education is an endogenous regressor, thus IV is chosen as the preferred method of analysis.

The motivation for the use of dummy variables to account for industry and occupation is because a part of the return to education may be affected by the kind of job/industry an individual works in; within certain industries, you may receive higher returns. Typically, one would expect the coefficient of interest (years in education) to fall once the occupation and industry are controlled for. This argument holds when comparing the Mincerian returns to education without the controls for industry and occupation to the education coefficient obtained from the full specification in Table 4. The IV estimates for education have reduced for both, men and women, after controlling for job characteristics – for men, an additional year of education brings about an 8.8 per cent increase in earnings after controlling for occupation and industry, whereas it was 12.1 per cent earlier; for women, the returns to education after controlling for occupation and industry is 10.4 per cent compared to 14.8 per cent without the controls. Women still receive a higher return to education relative to men. Looking at the additional industry and occupation dummy variables, it is clear which occupations pay more for men and women. Men receive relatively higher returns to employment in the armed forces, and high-skilled white-collar jobs such as managerial, professional and technical occupations. Women receive high returns to employment in the armed forces, professional and clerical jobs.

Table 5: OLS for agricultural and non-agricultural private and government workers

Dependent variable: log(earnings)	Agriculture	Private non-agriculture		Government non-agriculture	
Independent variables	(1)	(1)	(2)	(1)	(2)
Years in education	0.046***	0.067***	0.041***	0.086***	0.054***
	(6.85)	(32.95)	(18.26)	(29.39)	(13.75)
Age	0.218***	0.101***	0.083***	-0.012	-0.006
	(7.41)	(9.15)	(7.89)	(-0.37)	(-0.21)
Age^2	-0.004***	-0.002***	-0.001***	0.001	0.001
	(-6.56)	(-7.01)	(-5.87)	(1.47)	(1.28)
Age^3	$2.21x10^{-5}$***	$9.46x10^{-6}$***	$6.72x10^{-6}$***	$-1.35x10^{-5}$**	$-1.22x10^{-5}$**
	(5.37)	(4.66)	(3.49)	(-2.14)	(-1.95)
Female	-0.879***	-0.553***	-0.552***	-0.191***	-0.227***
	(-16.34)	(-37.79)	(-37.64)	(-11.47)	(-12.13)
Urban	-0.035	0.182***	0.138***	0.174***	0.122***
	(-0.32)	(13.25)	(10.28)	(10.42)	(7.49)
Married	-0.054	0.125***	0.118***	0.109***	0.102***
	(-0.83)	(8.06)	(7.81)	(4.48)	(4.26)
Sinhalese	0.197**	0.046***	0.038***	0.304***	0.141***
	(2.23)	(3.28)	(2.72)	(11.54)	(5.25)
Occupations:					
Managers			0.679***		0.003
			(16.74)		(0.04)
Professionals			0.407***		0.004
			(7.91)		(0.09)
Technicians			0.385***		-0.080*
			(11.03)		(-1.76)
Clerks			0.435***		-0.128***
			(10.60)		(-2.99)
Services			0.120***		-0.150***
			(3.46)		(-3.80)
Craft			0.029		-0.277***
			(0.91)		(-3.93)
Operators			0.206***		-0.123**
			(6.17)		(-2.24)
Elementary			-0.030		-0.288***
			(-1.12)		(-7.61)
Industries:					
Secondary			0.106***		0.437***
			(4.73)		(7.58)
Tertiary			-0.005		0.506***
			(-0.25)		(12.61)
Constant	5.069***	7.043***	7.376***	8.026***	8.175***
	(11.52)	(49.68)	(53.57)	(19.71)	(20.58)
N	4,420	15,123		3,871	

* indicates significance at the 10% level; ** 5% level; *** 1% level. The z-statistics are in parentheses. Default categories are: male, rural, not married, non-Sinhalese, armed forces, primary industry

Briefly exploring the first-stage IV estimation results controlling for occupation/industry presented in Appendix 1, the urban dummy for men becomes significant after imposing the controls, that is, men in urban regions are likely to attain more education compared to their rural counterparts. As one would expect, men and women employed in high-skilled white-collar jobs, especially as professionals, attain higher levels of education.

4.1.2 OLS versus IV: agriculture, non-agriculture private and government workers

In this section, individuals are disaggregated based on the type of employment – agricultural, non-agricultural private and government workers. Tables 5 and 6 present the earnings functions using OLS and IV respectively.

Firstly, we use the Mincerian specification given by (1) in Tables 5 and 6 to compare the OLS estimates of the returns to education to the IV estimates - the IV estimates are higher for all sub-groups. This suggests that the OLS estimates are under-estimating the returns to education. Comparing IV estimates across different sectors of employment, the returns to education are lowest for agricultural workers (8 per cent), and fairly similar for private and government workers (12.4 and 11.2 per cent respectively). These results, with the exception of the returns to education for private workers, are lower than the returns to education obtained previously for men (12.1 per cent) and women (14.8 per cent). One possible explanation could be that a part of the overall return to education might be getting individuals into higher-paying sectors.

Table 6: IV for agricultural and non-agricultural private and government workers

Dependent variable: log(earnings)	Agriculture	Private non-agriculture		Government non-agriculture	
Independent variables	(1)	(1)	(2)	(1)	(2)
Years in education	0.081***	0.124***	0.109***	0.112***	0.093***
	(5.97)	(28.30)	(17.78)	(20.06)	(8.29)
Age	0.222***	0.089***	0.080***	-0.029	-0.024
	(7.61)	(7.74)	(7.32)	(-0.93)	(-0.46)
Age^2	-0.004***	-0.001***	-0.001***	0.002**	0.001
	(-6.80)	(-5.32)	(-4.98)	(2.01)	(1.58)
Age^3	$2.3x10^{-5}$***	$6.49x10^{-6}$***	$5.39x10^{-6}$**	$-1.66x10^{-5}$**	$-1.40x10^{-5}$**
	(5.67)	(3.09)	(2.52)	(-2.59)	(-2.24)
Female	-0.806***	-0.525***	-0.525***	-0.215***	-0.214***
	(-14.81)	(-34.37)	(-34.18)	(12.55)	(-11.27)
Urban	-0.076	0.106***	0.102***	0.148***	0.123***
	(-0.70)	(7.02)	(7.24)	(8.41)	(7.36)
Married	0.016	0.142***	0.129***	0.123***	0.114***
	(0.24)	(8.70)	(8.06)	(4.93)	(4.67)
Sinhalese	0.114	-0.094***	-0.077***	0.175***	0.079**
	(1.24)	(-5.41)	(-4.50)	(5.40)	(2.53)

Occupations					
Managers			0.463***		-0.162*
			(10.15)		(-1.88)
Professionals			0.051		-0.184**
			(0.85)		(-2.53)
Technicians			0.182***		-0.207***
			(4.56)		(-3.45)
Clerks			0.144***		-0.236***
			(2.98)		(-4.27)
Services			-0.036		-0.200***
			(-0.92)		(-4.56)
Craft			-0.052		-0.303***
			(-1.48)		(-4.34)
Operators			0.081**		-0.155***
			(2.21)		(-2.81)
Elementary			-0.015		-0.268***
			(-0.53)		(-6.97)
Industries					
Secondary			0.013		0.361***
			(0.49)		(6.00)
Tertiary			-0.078***		0.392***
			(-3.26)		(8.10)
Constant	4.777***	6.716***	6.972***	8.111***	8.034***
	(10.63)	(45.00)	(47.23)	(19.72)	(20.07)
N	4,420	15,123		3,871	
Test for weak instruments	**Pass**	**Pass**	**Pass**	**Pass**	**Pass**
F-statistic:	722.56>10	1790.96>10	950.77>10	653.75>10	200.39>10
Test for over-identification	**Pass**	**Pass**	**Pass**	**Pass**	**Pass**
p-value:	0.76>0.05	0.69>0.05	0.94>0.05	0.95>0.05	0.95>0.05
Test for exogeneity	**Reject**	**Reject**	**Reject**	**Reject**	**Reject**
p-value:	0.00<0.05	0.00<0.05	0.00<0.05	0.00<0.05	0.00<0.05

* indicates significance at the 10% level; ** 5% level; *** 1% level. The z-statistics are in parentheses. Default categories are: male, rural, not married, non-Sinhalese, armed forces, primary industry

Figure 5: Age-Earnings profiles for agricultural and non-agricultural government and private sector workers

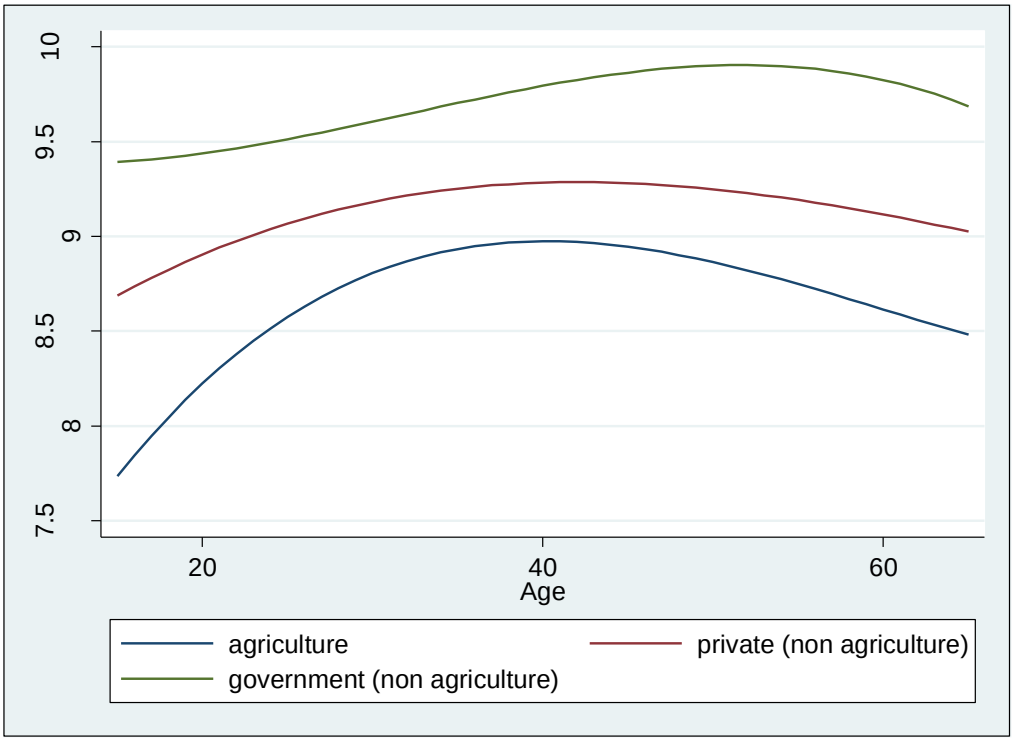


Figure 5 presents the age-earnings profiles for the three different sub-divisions using the IV estimates. Across the age distribution, the predicted earnings are highest for government sector workers and lowest for agricultural workers. For agricultural and private sector workers, the peak of earnings is when an individual is 40; for government workers, the peak of their earnings is just after 50. After these points, earnings are increasing at a decreasing rate.

The coefficient on the gender dummy variable indicates that women receive lower returns compared to men: on average 80 per cent lower in agriculture, 53 per cent less in the private sector and 22 per cent less in the public sector. As one would expect, individuals working in urban areas receive higher returns compared to their rural counterparts in both, the private (10.6 per cent more) and public non-agriculture (14.8 per cent more) sectors. Married individuals receive higher returns in the non-agriculture sectors compared to their unmarried counterparts. Individuals from a Sinhalese background receive lower returns than individuals from a non-Sinhalese background in the private sector; however the effect is reversed in the public sector – the returns are relatively higher for individuals from a Sinhalese background.

The first-stage IV results are presented in Appendix 1 which gives the relationship between the years of education and various individual characteristics such as age, gender, ethnicity and also the two instruments (change in the minimum years of schooling and

spouse's education level). Women in the public sector have higher levels of education compared to their male counterparts; however there are no gender differences in terms of education in the private sector. Individuals who reside in urban areas, unmarried or from a Sinhalese background attain more years of education. The coefficient on the policy instrument is positive for all three sub-groups implying that the policy change has had the positive impact on education that was expected. However, this impact is only significant at the 10 per cent significance level for people working in agricultural activities and the government sector. Possible explanations could be that individuals working in agricultural activities are likely to have lower levels of education, irrespective of whether the policy was implemented or not.

On the other hand, government workers, as it is seen from the descriptive statistics in Table 2, have the highest average years of education (11.3 years) compared to the other sub-groups (7.7 years in agriculture and 8.5 years in the private sector) – this could imply that these individuals already received the 9 years of compulsory schooling even before the policy was implemented, thus the fact that it came into effect did not have a significant impact. The instrument has a large positive impact (0.541) on the education levels of private sector workers. Spouse's education has a positive relationship with the individual's level of education. This can be explained by the fact that individuals get married to people with similar preferences/skills/knowledge.

Looking at the tests that were run to check for the validity of the use of IV (given at the end of the previous table), the IV estimates pass the tests. The tests for weak instruments and over-identification show that the instruments are not weak and are valid in all the three sub-groups. The test for exogeneity of the main regressor is rejected for the three categories of workers implying that education is endogenous and failing to treat it as an endogenous variable will lead to an inconsistent OLS estimate that is downward biased. Thus, IV was chosen as the preferred method of analysis to look at the wage functions while controlling for industry/occupation.

The full specifications are presented in Tables 5 and 6 by (2). These earnings functions include dummy variables to account for job and industry characteristics. The inclusion of occupation/industry controls has led to smaller coefficients on all the education variables. Focussing on the IV estimates in Table 6, an additional year of schooling will increase private sector earnings by 12.4 per cent without controlling for job characteristics; the coefficient drops to 10.9 per cent once characteristics are accounted for. Similarly for government workers, an additional year of schooling increases earnings by 11.2 per cent without job controls which drops to 9.2 per cent when the job controls are included. This suggests that perhaps a part of the overall effect is due to getting a better job, thus being able to receive higher returns to employment.

The coefficients on the occupation dummy variables indicate that managers, on average, receive higher returns in the private sector relative to other occupations. The industry dummies suggest that individuals in tertiary or service industries appear to receive

relatively lower returns compared to individuals working in manufacturing/secondary industries. One possible explanation for this could be the fact that a large proportion of private sector employment is informal. In the public sector, individuals working in the services/tertiary industries receive high returns relative to those working in the secondary industries.

The first stage IV estimates with the inclusion of occupation and industry dummies are available in Appendix 1. Some of the key findings will be discussed here. After controlling for job characteristics, the urban dummy is no longer significant for government workers, and the coefficient for private sector workers reduces from 0.93 to 0.39. The gender dummy is no longer positive suggesting that women in the private sector receive lower levels of education than men, once job characteristics are accounted for. As expected, individuals in white-collar jobs have higher levels of education in both sectors. Private sector workers in the secondary sector attain the highest level of education, on average, whereas government workers in the tertiary sector obtain the highest levels of education. Spouse's level of education has a positive effect on education even after controlling for occupation characteristics, whereas the policy variable has a positive impact on private sector workers' education levels, but not on government workers' education.

4.2 Returns to education by level (OLS)

As discussed in Section 2.1, there is a vast literature on the varying returns across the distribution of education (Psacharopoulos, 1994; Psacharopoulos and Patrinos, 2004; Colclough et al., 2010). Figure 1 presented the different types of education-earnings relationships – concave, convex or mixed. This section examines the earnings functions for individuals using levels of education; primary, (lower and upper) secondary and tertiary rather than using the total years of education. OLS estimation will be used instead of IV because there is a lack of instruments in order to carry out IV estimation if there are four endogenous regressors (for each level of education); an over-identified model cannot be formed. Table 7 gives the results for the pooled sample during the time period 2009/2010. We expect the coefficients to be lower than those reported if we estimated using IV.

The returns to education are increasing with every level of education; an individual with lower secondary education has returns 2.3 per cent higher relative to an individual with no education and an individual with upper secondary education has returns 61 per cent relative to an individual with no education. This convex relationship between education and earnings is in line with the findings by Riboud et al. (2007) and Colclough et al., (2010). However, the effects may vary across different sub-samples of the population and it is important to analyse sub-groups separately rather than assuming they have similar average characteristics. The next two sections consider the following sub-groups – (1)

men and women; (2) agriculture, private non-agriculture and government non-agriculture workers.

Table 7: OLS estimates for the pooled sample - education as a level variable

Dependent variable: log(earnings)	
Independent variables	Full sample
Primary (years)	0.056*
	(1.77)
Lower secondary (years)	0.023***
	(7.29)
Upper secondary (years)	0.606***
	(19.18)
Tertiary (0,1)	1.390***
	(33.11)
Age	0.115***
	(10.38)
Age^2	-0.002***
	(-7.69)
Age^3	$8.72x10^{-6}$***
	(4.58)
Female	-0.511***
	(-35.95)
Urban	0.255***
	(19.21)
Married	0.118***
	(7.28)
Sinhalese	0.060***
	(4.35)
Constant	6.887***
	(46.30)
N	23,414
R^2	0.21

* indicates significance at the 10% level; ** 5% level; *** 1% level. The z-statistics are in parentheses. Default categories are: no education, rural, not married, non-Sinhalese

4.2.1 Men versus women

First we estimate the returns to primary, secondary and tertiary levels of education, for men and women. Table 8 presents the earnings functions for the Mincerian and full specifications.

Table 8: OLS estimates for men and women

Dependent variable: log(earnings)	Men		Women	
Independent variables	(1)	(2)	(1)	(2)
Primary (years)	0.048***	0.036***	0.004	-0.006

	(5.52)	(4.19)	(0.46)	(-0.59)
Lower secondary (years)	0.046***	0.026***	0.006	0.011
	(6.86)	(3.96)	(0.53)	(0.93)
Upper secondary (years)	0.156***	0.097***	0.224***	0.099***
	(24.32)	(13.82)	(21.31)	(7.09)
Tertiary (0,1)	0.387***	0.289***	0.567***	0.448***
	(7.52)	(5.25)	(14.43)	(11.56)
Age	0.120***	0.109***	0.060***	0.062***
	(8.82)	(8.17)	(3.02)	(3.22)
Age^2	-0.002***	-0.002***	$9.38x10^{-4}$**	-0.001**
	(-6.57)	(-6.08)	(-2.03)	(-2.31)
Age^3	$9.02x10^{-6}$***	$8.40x10^{-6}$***	$2.94x10^{-6}$	$4.32x10^{-6}$
	(3.89)	(3.68)	(0.86)	(1.29)
Urban	0.223***	0.120***	0.229***	0.077***
	(14.94)	(8.32)	(8.71)	(2.99)
Married	0.239***	0.223***	-0.023	0.006
	(11.75)	(11.24)	(-0.91)	(0.24)
Sinhalese	0.086***	0.122***	-0.095***	-0.134***
	(5.27)	(7.75)	(-3.58)	(-4.83)
Occupations:				
Managers		0.842***		0.660***
		(18.49)		(8.60)
Professionals		0.678***		0.814***
		(13.11)		(12.22)
Technicians		0.554***		0.644***
		(14.15)		(10.35)
Clerks		0.497***		0.805***
		(11.29)		(12.25)
Services		0.373***		0.456***
		(9.73)		(6.99)
Craft		0.349***		0.020
		(9.34)		(0.36)
Operators		0.389***		0.465***
		(10.36)		(6.36)
Elementary		0.187***		0.224***
		(5.89)		(5.81)
Armed forces		0.949***		1.385***
		(20.49)		(12.33)
Industries:				
Secondary		0.148***		0.241***
		(5.87)		(5.52)
Tertiary		0.106***		0.151***
		(4.47)		(3.70)
Constant	6.586***	6.535***	7.493***	7.273***
	(36.27)	(36.27)	(28.58)	(28.16)
N	16,282		7,132	

* indicates significance at the 10% level; ** 5% level; *** 1% level. The z-statistics are in parentheses. Default categories are: no education, rural, not married, non-Sinhalese, agriculture, primary industry

First we explore the returns to education at different levels and across men and women using the Mincer earnings function (specification (1) in Table 8). Considering the men, education brings about higher returns at every level of education – an additional year of primary education increases earnings by 4.8 per cent whereas an additional year of upper secondary or tertiary education increases earnings by 15.6 and 38.7 per cent respectively. For women, lower education levels such as primary and lower secondary education are insignificant at the 10 per cent level, but the returns to higher levels of education such as upper secondary and tertiary (22.4 and 56.7 per cent, respectively) are higher than the returns that men receive at these levels (15.6 and 38.7 per cent, respectively).

Riboud et al. (2007) and Psacharopoulos and Patrinos (2004) reported similar results – the returns to primary education are higher for men compared to women; whereas the returns to upper secondary and tertiary education are higher for women. Looking at the coefficients for the education variables, the returns to education are increasing at every level of education, for both men and women. For women however, this increase from upper secondary to tertiary education is more substantial; women with a higher level of education are greatly rewarded. This could indicate that better access to higher levels of education for women could help reduce the gender wage-gap that exists in Sri Lanka and several other developing countries. The rest of the variables yield the expected coefficients; however, there are no significant differences at the 10 per cent level across married and unmarried women.

The inclusion of occupation and industry dummies in the earnings function (as done by specification (2) in Table 8) shows the types of jobs and industries that men and women are more likely to receive higher returns in. For men, occupations where they receive the highest returns include the armed forces, managerial and professional jobs. Men working in the secondary industry receive the highest returns (14.8 per cent higher than for those working in the primary industry) followed by the tertiary industry (10.6 per cent higher than for men working in the primary industry). Women working in the armed forces receive the highest returns, followed by those who work as professionals and clerks. The returns to female employment in the secondary industry are quite high (24 per cent higher than for primary industry employment) followed by the tertiary industry (15 per cent higher than primary industry employment).

The returns to education are lower at every level after including controls for job characteristics for both, men and women. The returns to upper secondary and tertiary education, for women, are 10 and 45 per cent respectively after including occupation/industry dummies whereas it was 22 and 57 per cent respectively without the controls. For men, the returns to upper secondary and tertiary education reduce to 10 and 29 per cent respectively. After controlling for occupation and industry, women receive similar returns to upper secondary education in comparison to men, and higher returns than men at the tertiary level. Such results may suggest that there could be tougher access to higher education for women compared to men and lower female labour force

participation (Riboud et al., 2007). However, women who overcome these two issues do relatively better than men in the labour market.

4.2.2 Agriculture, non-agriculture private and government workers

In this section we disaggregate the sample by type of employment – agricultural, non-agricultural private and government sector workers – in order to identify how the returns to education vary across the primary, secondary and tertiary levels. Table 9 presents the findings.

The estimates of the Mincerian earnings functions for the agricultural and non-agricultural private and government workers are given by specification (1). The coefficients on the levels of education are increasing for each sub-group at every level. Agricultural workers have the lowest returns at every level of education, apart from lower secondary where agricultural workers receive a return of 4.4 per cent compared to private sector workers receiving a return to lower secondary education of 3.7 per cent – a possible explanation for this could be the fact that the private sector pools formal and informal sector workers together, thus under (over)-estimating the returns to education for formal (informal) sector workers. Private sector workers receive the highest returns to upper secondary and tertiary education – an extra year of upper secondary (tertiary) education brings about a 14 (65) per cent increase in non-agriculture earnings. For public sector workers, the returns to education are 5 per cent at the primary level, increasing to 21 per cent at the tertiary level of education. The earnings function including controls for industry/occupation was estimated – given by specification (2) in Table 9. The coefficients for the returns to education are lower at every level after controlling for job characteristics; this is true for both, private and public sector workers.

Table 9: OLS estimates for agricultural and non-agricultural private and government workers

Dependent variable: log(earnings)	Agriculture	Private non-agriculture		Government non-agriculture	
Independent variables	(1)	(1)	(2)	(1)	(2)
Primary (years)	0.015	0.019***	0.008	0.050**	0.045*
	(0.78)	(3.27)	(1.42)	(2.09)	(1.88)
Lower secondary (years)	0.044**	0.037***	0.020***	0.100***	0.048**
	(2.31)	(6.77)	(3.75)	(8.24)	(3.87)
Upper secondary (years)	0.090***	0.135***	0.087***	0.115***	0.079***
	(3.30)	(21.41)	(12.76)	(15.65)	(9.35)

Tertiary (0,1)	-0.077	0.652***	0.612***	0.209***	0.170***
	(-0.22)	(10.64)	(9.99)	(9.88)	(7.78)
Age	0.216***	0.086***	0.073***	-0.017	-0.014
	(7.34)	(8.02)	(7.02)	(-0.54)	(-0.43)
Age^2	-0.004***	-0.002***	-0.001***	0.001	0.001
	(-6.49)	(-5.94)	(-5.06)	(1.58)	(1.46)
Age^3	$2.17x10^{-5}$***	$7.08x10^{-6}$***	$5.16x10^{-6}$***	$-1.42x10^{-5}$**	$-1.34x10^{-5}$**
	(5.29)	(3.60)	(2.71)	(-2.20)	(-2.09)
Female	-0.890***	-0.584***	-0.566***	-0.210***	-0.235***
	(-16.12)	(-40.24)	(-38.89)	(-11.26)	(-12.06)
Urban	-0.040	0.183***	0.142***	0.162***	0.119***
	(-0.36)	(13.53)	(10.65)	(9.78)	(7.29)
Married	-0.051	0.143***	0.130***	0.121***	0.110***
	(-0.79)	(9.33)	(8.66)	(5.08)	(4.70)
Sinhalese	0.190**	0.063***	0.044***	0.268***	0.133***
	(2.13)	(4.50)	(3.17)	(9.78)	(4.94)
Occupations:					
Managers			0.641***		-0.028
			(15.93)		(-0.39)
Professionals			0.258***		-0.033
			(5.00)		(-0.76)
Technicians			0.366***		-0.113**
			(10.19)		(-2.48)
Clerks			0.364***		-0.160***
			(8.65)		(-3.71)
Services			0.121***		-0.161***
			(3.50)		(-4.03)
Craft			0.046		-0.281***
			(1.42)		(-3.96)
Operators			0.223***		-0.123**
			(6.69)		(-2.20)
Elementary			-0.029		-0.282***
			(-1.07)		(-7.01)
Industries:					
Secondary			0.125***		0.440***
			(5.57)		(7.50)
Tertiary			0.014		0.505***
			(0.66)		(12.15)
Constant	5.208 ***	7.464***	7.651***	8.203***	8.305***
	(11.73)	(53.27)	(55.87)	(18.61)	(19.33)
N	4,420	15,123		3,871	

* indicates significance at the 10% level; ** 5% level; *** 1% level. The z-statistics are in parentheses. Default categories are: no education, male, rural, not married, non-Sinhalese, armed forces, primary industry

Considering the other independent variables in Table 9, the urban dummy is significant and positive for the non-agriculture sub-groups suggesting than individuals working in private and public sector jobs in urban areas receive higher returns to employment relative to their rural counterparts. The gender coefficient is negative – women receive

lower returns than men, especially in agriculture and the private sector. The coefficient, although still significant, is smaller in the public sector. Riboud et al. (2007) found that gender-related wage gaps between men and women are around 50 per cent in Sri Lanka. There could be many possible reasons for the presence of this gap such as type of employment, sector of work and discrimination.

4.3 Production function at the household level

In this section, we estimate the production function for agricultural and non-agricultural households. Here we focus on the self-employed in agricultural and non-agricultural activities. Non-agriculture is defined as activities such as construction, trade, mining and quarrying, manufacturing, transport, and services such as restaurants, hotels, etc. The main variable of interest is the natural logarithm of output generated from the entire enterprise/household, rather than individual earnings. As discussed in Section 3.2, an estimation at the household level allows us to incorporate unpaid family workers into the analysis – who may contribute to output, but will be dropped out of the individual-level earnings function as they do not report any form of income. Table 10 presents the production function for agricultural and non-agricultural households in self-employment.

The results suggest that input, land and labour have a significant (at the 1 per cent level) and positive association with agricultural output, as expected. For non-agricultural households, input and labour have a positive and significant association with output at the 1 per cent level[23]. As explained in Section 3.2, the factors of production (inputs, land, and labour) are scaled; the coefficients for the respective factors seen in Table 10 are output elasticities evaluated at the means. These elasticities imply that if labour were doubled, agricultural output would increase by 11 per cent. Appleton and Balihuta (1996) found similar values in Uganda. If the land holdings were doubled, agricultural output would increase by almost 25 per cent, implying that agriculture seems to be more land-constrained than labour-constrained. Although this may not be true for certain areas, it holds on aggregate.

Likelihood ratio tests were used to check if it is reasonable to restrict the translog function to a Cobb-Douglas function. Table 11 presents the results. The likelihood ratio test estimates two models, after which a comparison is made. This test compares the log likelihoods of the models, and then tests whether the difference is statistically different. If

[23] Jones (2001) and Appleton and Balihuta (1996) observed an insignificant and small coefficient on labour, suggesting that this could be due to the use of the number of individuals working in the household rather than using the hours of work as a key variable to measure labour. The aforementioned papers explained that, if the number of workers employed in non-agricultural activities are doubled, that might not necessarily imply that the production or hours worked will double as a result of it. However, if an individual spends twice the number of hours as he did previously, consequently, the productivity may have also doubled. This measurement error could lead to a downward bias of the coefficient to zero. Unfortunately, the survey used in this study does not give information on the hours worked by individuals in order to use this as a measure of labour in the production function. The coefficient on labour however does not indicate a downward bias towards zero.

there is a significant difference, the less restrictive model (with more variables; in this case, the translog function) fits the data significantly better than the more restrictive one (the Cobb-Douglas function). Leaving out the second-order terms, as the Cobb-Douglas function would do, will significantly reduce the fit of the model.

Table 10: Production function for self-employed (household level)

Dependent variable: log(output)	Agriculture		Non-agriculture	
Independent variables	Coefficient	T-ratio	Coefficient	T-ratio
log(input)	0.712***	54.03	0.847***	85.81
log(land)	0.235***	14.66		
log(labour)	0.108***	3.80	0.158***	5.60
log input^2	0.027***	4.57	0.035***	14.77
log land^2	0.039***	4.08		
log labour^2	0.112*	1.93	0.117***	2.74
log(input)*log(land)	-0.057***	-4.34		
log(input)*log(labour)	-0.008	-0.37	0.060***	4.68
log(land)*log(labour)	-0.031	-0.96		
No education (proportion)	-0.015	-0.15	0.161	1.61
Primary (average years)	-0.001	-0.06	0.021	1.10
Lower Secondary (average years)	0.005	0.43	0.014	1.57
Upper Secondary (average years)	-0.039***	-3.47	0.032***	3.50
Tertiary (proportion)	-0.094	-1.32	0.210***	3.20
Age (average)	0.011	0.44	-0.014	-0.91
Age^2 (average)	-1.57×10^{-4}	-0.30	4.40×10^{-4}	1.29
Age^3 (average)	4.74×10^{-7}	0.13	-4.34×10^{-6}*	-1.76
Female (proportion)	0.004	0.09	-0.331***	-10.50
Rural (0,1)	0.119**	2.03	-0.076***	-4.46
Married (proportion)	-0.004	-0.10	0.077***	2.72
Sinhalese (proportion)	-0.032	-0.60	0.038*	1.84
Constant	11.099***	28.96	12.039***	56.09
Number of households	3,460		3,320	
R^2	0.84		0.88	

* indicates significance at the 10% level; ** 5% level; *** 1% level; Default categories: male, urban, not married, non-Sinhalese; Inputs include spending on raw materials and capital; Primary, lower and upper secondary education are measured as the average number of years spent in that particular level of education by all working persons in the household; no education (tertiary education) is measured as the proportion of working persons who have not gone to school (completed tertiary education).

Table 11: Likelihood ratio tests

	Agriculture	**Non agriculture**
Likelihood ratio test	LR chi2 (6) = 76.26 Prob > chi2 = 0.000	LR chi2 (3) = 531.55 Prob > chi2 = 0.000

Having run this test for both agriculture and non-agriculture, the results were as follows. For the agricultural households, the chi-squared value for the test was 75.31, with a p-value for the chi-squared of zero with six degrees of freedom (since 6 parameters have been removed from the model). For the non-agricultural households, the chi-squared value for the test was 510.06, with a p-value for the chi-squared of zero with three degrees of freedom (since 3 parameters have been removed). Hence, the Cobb-Douglas function can be rejected in favour of the translog function.

Looking at the rest of the input variables, the squared input, land and labour coefficients are positive and highly significant (apart from the squared-labour term for agriculture which is significant at the 10 per cent level). This implies that output, both agricultural and non-agricultural, is increasing at an increasing rate with the inputs, land and labour. Moving on to the interaction terms, the interaction between agricultural input and land has a negative relationship with agricultural output. On the other hand, the interaction between non-agricultural input and labour has a positive relationship with non-agricultural output.

The coefficients on the education variables can be analysed as follows. Firstly looking at agricultural households, the average years of upper secondary education obtained by the workers involved in agricultural activities has a negative association with output. Negative coefficients on some of the education variables have been observed in the meta-analysis by Phillips (1994). One interpretation of the negative coefficients is that educated workers would choose to spend more time on another occupation, spending less hours working on the farm[24] - that is, they have more opportunities in the labour market and choose to lower their productivity in agriculture. Appleton and Balihuta (1996) observed a positive relationship between primary education and agricultural output for Uganda, however in the case of Sri Lanka, primary education is not significant (at the 10 per cent level) in affecting agricultural productivity. In general, several studies reviewed in Appleton and Balihuta (1996) and Lockheed et al. (1980) showed that more educated agricultural workers are not necessarily more productive than uneducated agricultural workers in developing countries. This could be the reason why tertiary and lower secondary education are insignificant in the production function for agricultural households. These insignificant results on the education variables can be attributed to the low levels of technology being used in rural labour markets (Jones, 2001). On the other hand for non-agricultural households, the results suggest that the average years of upper secondary education obtained by the employed in the household are associated with higher levels of output by 3.2 per cent whereas tertiary education is associated with higher levels of output by 21 per cent. Focussing on the demographic characteristics in the production function, the proportion of working women in the household is negatively associated with non-agricultural productivity. Not surprisingly households in rural areas

[24] If the data gave any information on the hours worked in the farm, it would have been useful in obtaining an estimate showing the relationship between education and farm productivity.

report higher levels of agricultural productivity compared to those from urban areas, whereas the opposite is true for non-agricultural households.

5. Conclusion

This paper explored the rates of returns to education in Sri Lanka for the year 2009/10. It contributes to existing literature by accounting for the endogeneity of education which has not been dealt with in previous studies for the case of Sri Lanka. To estimate the returns to education, three approaches were used. Firstly, an instrumental variables approach was used alongside a comparable OLS estimation to account for the endogeneity bias. Spouse's education and the implementation of compulsory schooling until the age of 14 in 1997 were used as instruments to deal with the potential endogeneity of education. The Hausman test, which tested the null hypothesis of the exogeneity of education was rejected, suggesting that education is endogenous. Having run tests to ensure that the instruments used in the IV estimation were valid, IV was chosen as the preferred method of analysis.

The findings suggested that the years of education bring about an 8 per cent return to education for agricultural workers, a 12.4 per cent return for private non-agricultural workers, and an 11.2 per cent return for government non-agricultural workers, which is higher than the corresponding OLS estimates. This is in line with previous studies that showed a downward bias in the OLS estimates. The IV analysis of the male and female workers suggested that female workers receive a higher return to education of 14.8 per cent compared to their male counterparts (12.1 per cent).

To examine how the returns to education vary as an individual obtains more education, OLS was used for the same sub-groups to identify differences by primary, secondary and tertiary education. The results for the agricultural and non-agricultural workers suggested that the rates of returns are increasing with the level of education, with tertiary education bringing about the highest returns. Individuals working in the government sector receive higher returns to education at the primary and lower secondary levels while the private sector pays higher returns at the upper secondary and tertiary levels. The analysis of the men and women also showed increasing returns, with women receiving higher returns to upper secondary and tertiary education in comparison to men.

Finally, the translog production function was estimated at the household level using OLS for agricultural and non-agricultural households involved in self-employment. The presence of individuals with upper secondary education had a negative association with agricultural output; whereas for non-agricultural households, productivity was positively associated with the presence of individuals with upper secondary and tertiary education.

Overall, the results from this study imply that primary education has relatively lower returns compared to higher levels of education. The high returns to post-primary

education suggest that this type of education is a potential area for further investment, if the government wishes to focus on education policies that improve equity and efficiency of higher education. Generally, the results are consistent with previous studies but there is scope for further research. The returns to education were measured using individual incomes/wages. This does not take into account the positive externalities of primary education including both economic and non-market benefits (for example, lower crime rates, better environment, low fertility and mortality rates). However, it is difficult to assign monetary values to these non-market benefits which are taken into account by policy makers for further investment.

Due to a lack of data on the level of public expenditure on each education level, it is difficult to account for this using the Mincerian returns. There are certain other limitations to this study that have been identified. Firstly, people may have more than one occupation, leading to a bias in the way their wage profiles have been created. Secondly, the survey does not record the number of hours worked by an individual. Farmers, for example, work less during certain seasons, and more during others. The survey does not give a breakdown of the hourly wage, which leads to an imperfect measure of labour in the production function and the wages in the earnings function. Finally, this paper does not account for the non-random way in which individuals choose into various sectors of employment.

References

Adelman, I., 1961, *Theories of Economic Growth and Development*, Stanford, CA: Stanford University Press

Angrist, J. D. and Krueger, A. B., 1991, "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106 (4), pp. 979 - 1014

Appleton, S. and Balihuta, A., 1996, "Education and Agricultural Productivity: Evidence from Uganda", *Journal of International Development*, 8 (3), pp. 415 - 444

Appleton, S., 2001, "Education, Incomes and Poverty in Uganda in the 1990s." CREDIT Research Paper No. 01/22

Arrow, K. J., 1973, "Higher Education as a Filter", *Journal of Public Economics*, 2 (3), pp. 193 – 216

Ashenfelter, O., Harmon, C. and Oosterbeek, H., 1999, "A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias", *Labour Economics*, 6 (4), pp. 453 - 470

Aturupane, H., 2011, "Transforming School Education in Sri Lanka: From cut stones to polished jewels", *World Bank*, Report No: 66036

Awan, M. S., Malik, N., Sarwar, H. and Waqas, M., 2011, "Impact of Education on Poverty Reduction". *International Journal of Academic Research*, 3 (1), pp. 659 – 664

Becker, G. S., 1964, *Human Capital*, New York: Columbia University Press

Becker, G. S., 1994, *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, 3rd Edition. Chicago: The University of Chicago Press

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

Blackburn, M. and Neumark, D., 1993, "Omitted-ability bias and the increase in the return to schooling", *Journal of Labor Economics*, 11 (3), pp. 521 – 544

Blackburn, M. and Neumark, D., 1995, "Are OLS estimates of the return to schooling biased downward? Another look", *the Review of Economics and Statistics*, 77 (2), pp. 217 – 230

Bound, J., Jaeger, D., and Baker, R., 1995, "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variables is Weak", *Journal of the American Statistical Association*, 90 (430), pp. 443 – 450

Brown, P. H. and Park, A., 2002, "Education and Poverty in Rural China", *Economics of Education Review*, 21, pp. 523 – 541

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

Card D., 2001, "The Causal Effect of Education", Mimeo, Center for Labor Economics, University of California, Berkley

Carnoy, M., 1994, *Faded Dreams: The Politics and Economics of Race in America*, Cambridge University Press

Colclough, C., Kingdon, G. and Patrinos, H., 2010, "The Changing Pattern of Wage Returns to Education and its Implications", *Development Policy Review*, 28 (6), pp. 733 - 747

Connell, R. W., 1994, **"**Poverty and Education", *Harvard Educational Review*, 64 (2), pp. 125 – 149

Datt, G. and Gunerwardena, D., 1997, "Some Aspects of Poverty in Sri Lanka: 1985-90", *World Bank*, Policy Research Working Paper 1738

Department of Census and Statistics, Available: http://www.statistics.gov.lk/. Last accessed: 28th September 2016

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

Gunerwardena, D., 2002, "Reducing the Gender Wage Gap in Sri Lanka: Is Education Enough?" *Sri Lanka Economic Journal*, 3(2), pp. 57 - 103

Gwartney, J. and Stroup, R., 1973, "Measurement of Employment Discrimination According to Sex", *Southern Economic Journal,* 39 (4), pp. 575 - 587

Harmon C. and Walker I., 1995, "Estimates of the Economic Return to Schooling for the United Kingdom", *American Economic Review*, 85 (5), pp. 1278 - 1286

Jacoby, H. G., 1992, "Productivity of Men and Women and the Sexual Division of Labour in Peasant Agriculture of the Peruvian Sierra", *Journal of Development Economics*, 37 (1 & 2), pp. 265 – 287

Jones, P. 2001, "Are educated workers really more productive?" *Journal of Development Economics*, 64 (1), pp. 57 - 79

Kingdon, G., Patrinos, H. A., Sakellariou, C. and Söderbom, M., 2008, "International Pattern of Returns to Education", Washington, DC: World Bank (mimeo)

Krueger, A. B., Lindahl, M., 2001, "Education for Growth: Why and For Whom?" *Journal of Economic Literature,* 39 (4), pp. 1101 – 1136

Kuznets, S., 1956, "Economic Growth and Income Inequality", *American Economic Review*, 45 (1), pp. 1 – 28

Li, H., 2003, "Economic Transition and Returns to Education in China", *Economics of Education Review*, 22 (3), pp. 317-28

Lipton, M. and Ravallion, M., 1995, "Poverty and Policy", In: Behrman J. and Srinivasan, T.N. *Handbook of Development Economics*, Volume 3. Amsterdam: Elsevier Science. pp. 2551 - 2657

Mincer, J., 1974, *Schooling, Experience, and Earnings*, Columbia University Press, New York

Moav, O., 2004, "Cheap Children and the Persistence of Poverty", *the Economic Journal*, 115 (500), pp. 88 - 110

OECD, Education at a Glance, 2006, Paris, www.oecd.org/document/52/0,2340,en_2649_34515_37328564_1_1_1_1,00.html

Parker, S. C. and Van Praag, C. M., 2006, "Schooling, capital constraints, and entrepreneurial performance: The endogenous triangle", *Journal of Business & Economic Statistics*, 24 (4), pp. 416 – 431

Pencavel, J., 1998, "Assortative mating by schooling and the work behaviour of wives and husbands", *American Economic Review*, 88 (2), pp. 326 – 329

Phillips, J. M., 1994, "Farmer Education and Farmer Efficiency: A Meta-Analysis", *Economic Development and Cultural Change*, 43 (1), pp. 149 – 165

Psacharopoulos, G. and Patrinos, H., 2004, "Returns to Investment in Education: A Further Update", *Education Economics*, 12 (2), pp. 111 - 134

Psacharopoulos, G., 1994, "Returns to investment in education: a global update" *World Development*, 22 (9), pp. 1325 - 1344

Psacharopoulos, G., 1988, "Education and Development: A Review", *World Bank Research Observer*, 3 (1), pp. 99-116

Ranasinghe, A. and Hartog, J., 2002, "Free-education in Sri Lanka: Does it eliminate the family effect?" *Economics of Education Review*, 21 (6), pp. 623 - 633

Ranasinghe, A. and Hartog, J., 1997, "Investment in Post-Compulsory Education in Sri Lanka", Tinbergen Institute Discussion Paper 97-021/3

Riboud, M., Savchenko, Y. and Tan, H., 2007, "The Knowledge Economy and Education and Training in South Asia". Washington, DC: World Bank

Schultz, T. P., 1961, "Investments in Human Capital", *American Economic Review*, 51 (1), pp. 1 – 17

Schultz, T. P., 2004, "Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education", *Journal of African Economies,* 13 (2), pp. 95-148

Staiger, D. and Stock, J., 1997, "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 65 (3), pp. 557 – 586

Stevens, P. and Weale, M., 2007, "Education and economic growth", In: Johnes, G. and Johnes, J. *International handbook on the economics of education*. Massachusetts: Edward Elgar Publishing. pp. 164 – 188

Temple, J., 2001, "Growth effects of education and social capital in the OECD countries". Centre for Economic Policy Research

Trostel, P., Walker, I. and Wooley, P., 2002, "Estimates of the economic return to schooling for 28 countries", *Labour Economics*, 9 (1), pp. 1–16

Vila, L., 2000, "The Non-Monetary Benefits of Education", *European Journal of Education*, 35 (1), pp. 21 – 32

World Bank, 2014, Available: http://data.worldbank.org/. Last accessed: 28th October 2014

World Bank, 2005, "Attaining the Millennium Development Goals in Sri Lanka", Colombo: World Bank

World Bank, 2001, "World Development Report 2000/2001. Attacking Poverty: Overview", Washington, D.C.

Appendix 1: First stage IV regression results

First-stage IV estimates for men and women

Dependent variable: years in education	Men		Women	
Independent variables	(1)	(2)	(1)	(2)
Age	0.040	-0.015	0.413***	0.167*
	(0.74)	(-0.31)	(4.57)	(1.92)
Age^2	-0.001	$-2.06x10^{-4}$	-0.010***	-0.005**
	(-1.05)	(-0.20)	(-5.24)	(-2.54)
Age^3	$6.67x10^{-6}$	$1.62x10^{-6}$	$5.83x10^{-5}$***	$2.61x10^{-5}$**
	(0.90)	(0.24)	(4.66)	(2.17)
Urban	0.819	0.330***	1.096***	0.336***
	(15.83)	(6.84)	(11.94)	(4.08)
Married	-5.701***	-4.393***	-5.244***	-3.326***
	(-50.71)	(-40.32)	(-34.96)	(-22.51)
Sinhalese	0.913***	0.892***	2.935***	2.319***
	(15.16)	(15.92)	(28.33)	(24.44)
Occupations:				
Managers		1.287***		1.491***
		(10.73)		(6.63)
Professionals		3.685***		4.095***
		(24.60)		(21.19)
Technicians		1.375***		2.160***
		(12.82)		(10.60)
Clerks		2.277***		2.957***
		(18.56)		(16.01)
Services		0.736***		1.258***
		(6.41)		(5.69)
Craft		0.161		0.584***
		(-1.52)		(2.99)
Operators		0.385***		0.457*
		(3.63)		(1.97)
Elementary		-0.852***		-0.982***
		(-9.48)		(-6.80)
Armed forces		0.849***		2.011***
		(4.03)		(3.87)
Industries				
Secondary		0.798***		0.664***
		(9.35)		(3.86)
Tertiary		0.827***		0.554***
		(10.41)		(3.40)
Policy	0.326***	0.343***	0.920***	0.316*
	(2.75)	(3.13)	(4.46)	(1.69)
Spouse's education	0.607***	0.469***	0.635***	0.403***
	(75.89)	(56.26)	(50.63)	(31.37)

Constant	7.966***	8.074***	2.125	4.802***
	(9.72)	(10.64)	(1.54)	(3.59)
N	16,282		7,132	

* indicates significance at the 10% level; ** 5% level; *** 1% level. The t-statistics are in parentheses. Default categories are: rural, not married, non-Sinhalese, agriculture, primary industry. Controlled for missing values of spouse's level of education

First-stage IV estimates for agricultural and non-agricultural private and government workers

Dependent variable: years in education	Agriculture	Private non-agriculture		Government non-agriculture	
Independent variables	(1)	(1)	(2)	(1)	(2)
Age	-0.081	0.179***	0.054	0.669***	0.286*
	(-0.62)	(3.01)	(1.07)	(3.10)	(1.58)
Age^2	0.001	-0.005***	-0.002**	-0.016***	-0.007*
	(0.56)	(-4.17)	(-2.17)	(-2.98)	(-1.72)
Age^3	$-1.07x10^{-5}$	$4.07x10^{-5}$***	$1.76x10^{-5}$**	$1.13x10^{-4}$***	$5.58x10^{-5}$*
	(-0.71)	(4.35)	(2.24)	(2.72)	(1.66)
Female	0.246**	-0.026	-0.104**	0.757***	-0.132*
	(2.08)	(-0.47)	(-2.02)	(8.24)	(-1.57)
Urban	0.877***	0.934***	0.394***	0.594***	-0.020
	(4.24)	(17.49)	(8.12)	(6.31)	(-0.25)
Married	-4.016***	-4.666***	-3.379***	-7.076***	-3.828***
	(-20.99)	(-42.59)	(-31.78)	(-26.16)	(-16.34)
Sinhalese	1.080***	1.491***	1.138***	2.765***	1.110***
	(5.93)	(24.73)	(20.64)	(16.20)	(7.51)
Occupations:					
Managers			2.352***		3.709***
			(16.45)		(12.58)
Professionals			4.303***		4.187***
			(25.34)		(21.02)
Technicians			2.329***		2.915***
			(17.49)		(14.66)
Clerks			3.688***		2.502***
			(25.76)		(12.90)
Services			1.776***		1.271***
			(12.58)		(6.42)
Craft			0.924***		0.853***
			(7.23)		(3.14)
Operators			1.358***		0.896***
			(10.30)		(3.89)
Elementary			-0.121		-0.084
			(-1.05)		(-0.43)

Industries					
Secondary			1.114***		1.294***
			(12.59)		(4.77)
Tertiary			0.830***		2.277***
			(9.67)		(11.27)
Policy	0.657*	0.541***	0.384***	0.506*	0.266
	(1.78)	(4.57)	(3.76)	(1.79)	(1.13)
Spouse's education	0.546***	0.544***	0.401***	0.630***	0.339***
	(38.01)	(59.72)	(43.51)	(36.12)	(19.96)
Constant	8.125***	6.114***	6.140***	-0.121	3.446
	(3.73)	(7.10)	(8.26)	(-0.04)	(1.39)
N	4,420	15,123		3,871	

* indicates significance at the 10% level; ** 5% level; *** 1% level. The t-statistics are in parentheses. Default categories are: male, rural, not married, non-Sinhalese, armed forces, primary industry. Controlled for missing values of spouse's level of education

Appendix 2: Other instruments

IV estimation for the full sample using quarter of birth as an instrument

	First-stage regression		**IV regression**	
Dependent variable	Years in education		log (earnings)	
	Coefficient	t-statistic	Coefficient	z-statistic
Independent variables				
Years in Education			0.149**	2.28
Age	0.174***	4.62	0.107***	6.79
Age^2	-0.005***	-6.04	-0.002***	-4.07
Age^3	$3.23x10^{-5}$***	5.37	$6.99x10^{-6}$**	2.49
Female	0.162***	3.08	-0.505***	-28.29
Urban	1.565***	28.76	0.129	1.26
Married	0.109*	1.76	0.110***	6.05
Sinhalese	2.810***	48.36	-0.162	-0.88
Instrument: quarter of birth (base: quarter 1)				
Quarter 2	-0.119*	-1.86		
Quarter 3	0.140**	2.18		
Quarter 4	0.084	1.32		
Constant	5.533***	10.89	6.188***	15.70
N	23,414			
Test for weak instruments	**Does not pass**			
F-statistic:	5.89<10			
Test for over-identification	**Pass**			
p-value:	0.13>0.05			
Test for exogeneity	**Do not reject**			
p-value:	0.31>0.05			

* indicates significance at the 10% level; ** 5% level; *** 1% level; Default categories are: male, rural, not married, non-Sinhalese

IV estimation for the full sample using parents' education as an instrument

	Reduced form		**IV regression**	
Dependent variable	Years in education		log (earnings)	
	Coefficient	t-statistic	Coefficient	z-statistic
Independent variables				
Years in Education			0.165***	17.43
Age	0.169***	4.64	0.105***	8.96
Age^2	-0.004***	-5.30	-0.002***	-6.16
Age^3	$2.46x10^{-5}$***	4.27	$7.08x10^{-6}$***	3.53
Female	0.194***	3.73	-0.513***	-34.80
Urban	1.427***	26.73	0.102***	5.02
Married	0.540***	8.51	0.078***	4.18
Sinhalese	2.532***	43.73	-0.206***	-6.79
Instrument: parents' education	0.220***	19.73		
Mother's education (years)	0.184***	12.83		
Father's education (years)				
Constant	3.160***	6.48	6.082***	39.32
N	23,414			
Test for weak instruments	**Pass**			
F-statistic:	504.47>10			
Test for over-identification	**Pass**			
p-value:	0.17>0.05			
Test for exogeneity	**Reject**			
p-value:	0.00<0.05			

* indicates significance at the 10% level; ** 5% level; *** 1% level; Default categories are: male, rural, not married, non-Sinhalese; Controlled for missing values of parents' level of education

Appendix 3: Variables used in the earnings and production functions

Variable	Description	Number of observations	Mean	Minimum	Maximum	Standard deviation
Variables used in the earnings function (at the individual level):						
log(earned income)	natural log of all earned income	23,414	9.18	0	17.42	1.05
education (years)	total years of education	23,414	8.79	0	19	3.84
age	individual's age	23,414	41.71	15	65	13.47
female	female=1 if the individual is a woman, else zero (man)	23,414	0.30	0	1	0.46
urban	urban=1 if the individual is from an urban area, else zero (rural)	23,414	0.24	0	1	0.42
married	married=1 if the individual is reported to be married, else zero (unmarried/divorced/widowed/separated)	23,414	0.75	0	1	0.44
Sinhalese	Sinhalese=1 if the individual is from a Sinhalese ethnic background, else zero (Muslim/Tamil/Burgher/Other)	23,414	0.78	0	1	0.41
policy	change in compulsory schooling to 14 years in 1997 (policy=1 if the individual is in school from 1997 onwards, else zero)	23,414	0.14	0	1	0.35
primary	years of primary schooling	23,414	4.56	0	5	1.19
lower secondary	years of lower secondary schooling	23,414	2.81	0	4	1.67
upper secondary	years of upper secondary schooling	23,414	1.27	0	4	1.48
tertiary	tertiary=1 if individual has completed tertiary education, else zero (no education/primary/secondary)	23,414	0.04	0	1	0.19
no education	no education=1 if individual has not attended school, else zero (attained any level of education)	23,414	0.04	0	1	0.19
manager	manager=1 is individual is working as manager/ senior professional/ legislator	23,414	0.07	0	1	0.24
professional	professional=1 if individual is working as professional	23,414	0.07	0	1	0.25
technician	technician=1 if individual is working as technician/ associate professional	23,414	0.10	0	1	0.31
clerk	clerk=1 if individual is working as clerk/ clerical support	23,414	0.04	0	1	0.20
service	service=1 if individual is working in services/ sales	23,414	0.07	0	1	0.25

craft	craft=1 is individual is working in craft and related trades	23,414	0.14	0	1	0.35
operator	operator=1 if individual is working as plant and machine operator/ assembler	23,414	0.07	0	1	0.26
elementary	elementary=1 if individual is working in elementary occupation	23,414	0.23	0	1	0.42
armed forces	armed=1 if individual is working in the armed forces	23,414	0.01	0	1	0.06
agriculture	agriculture=1 if individual is working as agriculture/ forestry/ fishery worker	23,414	0.20	0	1	0.40
primary industry	primary industry=1 if individual is working in agriculture/ forestry/ mining/ quarrying/ hunting	23,414	0.30	0	1	0.46
secondary industry	secondary industry=1 if individual is working in manufacturing/ electricity, gas and water supply/ construction/ wholesale and retail trade	23,414	0.22	0	1	0.42
tertiary industry	tertiary industry=1 if individual is working in hotels and restaurants/ real estate, renting and business activities/ public administration and defence/ education/ health, social and community work	23,414	0.47	0	1	0.50
Variables used in the production function for agriculture (at the household level):						
agricultural output	sum of agricultural output (value in rupees)	3,460	92,488	100	2,240,000	148,056
agricultural input	sum of agricultural input (value in rupees)	3,460	31,959	25	1,500,500	60,556
land	sum of agricultural land (value in acres)	3,460	1.95	0.006	84	3.12
labour	sum of employed persons in household	3,460	1.71	1	6	0.81
average age	average age of employed persons	3,460	44.27	17	63	10.67
women (proportion)	Proportion of employed women	3,460	0.29	0	1	0.32
married (proportion)	Proportion of married employed persons	3,460	0.80	0	1	0.35
Sinhalese (proportion)	Proportion of Sinhalese employed persons	3,460	0.95	0	1	0.22
primary (average)	average years of primary schooling acquired by employed persons	3,460	4.57	0	5	1.00
lower secondary (average)	average years of lower secondary schooling acquired by employed persons	3,460	2.77	0	4	1.47
upper secondary	average years of upper secondary schooling acquired by employed	3,460	1.14	0	4	1.27

(average)	persons					
tertiary (proportion)	proportion of employed persons with tertiary education	3,460	0.03	0	1	0.15
no education (proportion)	proportion of employed persons who have not attended school	3,460	0.03	0	1	0.14
rural	rural=1 if household is located in a rural area, else zero (urban)	3,460	0.97	0	1	0.18
Variables used in the production function for non-agriculture (at the household level):						
non-agriculture output	sum of non-agricultural output (value in rupees)	3,320	223,073	180	180,000,000	3,757,459
non-agriculture input	sum of non-agricultural input (value in rupees)	3,320	179,744	25	144,000,000	3,188,833
labour	sum of employed persons in household	3,320	1.73	1	7	0.86
average age	average age of employed persons	3,320	43.37	20	66	9.92
women (proportion)	Proportion of employed women	3,320	0.26	0	1	0.32
married (proportion)	Proportion of married employed persons	3,320	0.80	0	1	0.34
Sinhalese (prop)	Proportion of Sinhalese employed persons	3,320	0.79	0	1	0.40
primary (average)	average years of primary schooling acquired by employed persons	3,320	4.77	0	5	0.74
lower secondary (average)	average years of lower secondary schooling acquired by employed persons	3,320	3.15	0	4	1.29
upper secondary (average)	average years of upper secondary schooling acquired by employed persons	3,320	0.34	0	4	1.24
tertiary (proportion)	proportion of employed persons with tertiary education	3,320	0.02	0	1	0.12
no education (proportion)	proportion of employed persons who have not attended school	3,320	0.02	0	1	0.11
rural	rural=1 if household is located in a rural area, else zero (urban)	3,320	0.67	0	1	0.47