Education, Sectoral Choice and the Urban-Rural Welfare Gap in Sri Lanka

Nirodha Anuththari Bandara, BA (Hons), MSc

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

October, 2016

Abstract

This thesis is centred on three core issues in the context of Sri Lanka - the rates of returns to education across individuals in different types of employment, the contributors towards the urban-rural welfare gap for the years 2002 and 2009/10, and finally the determinants of employment and earnings across sectors.

First, the rates of returns to education are analysed using an Instrumental Variables approach in order to address the endogeneity bias associated with measuring education. Secondly, the returns to education are estimated at different levels; a convex relationship between education and earnings is observed. Finally, a production function is estimated for agricultural and non-agricultural self-employed households. Higher education shares a positive relationship with non-agricultural output, but a negative relationship with agricultural output.

In the second part of the thesis, we first identify the determinants of urban and rural expenditure using an unconditional quantile regression. Next, we examine the urban-rural welfare gap in 2002 and 2009/10 using a variant of the threefold Blinder-Oaxaca decomposition; identifying the characteristics and returns to characteristics that contribute towards the welfare gap in both years, and across the expenditure distribution. At a given point in time, the welfare gap is larger between richer urban and rural households relative to poorer households. We find the gap to have fallen considerably between 2002 and 2009/10.

The final part in the thesis examines the sectoral choices and earnings. The labour market is disaggregated into 5 sectors. This chapter controls for two forms of possible bias – sample selectivity and endogeneity of education in earnings. The determinants of sector choice are analysed using a multinomial logit. We observe that individuals with the highest levels of education get in to the public and formal private sectors, whereas the least educated are likely to join the informal and agricultural sectors. The earnings functions suggest that the returns to education vary greatly across the sectors.

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my supervisors, Doctor Simon Appleton and Doctor Trudy Owens for their continuous support and helpful critique throughout the progression of this thesis. I am forever grateful to have learnt under your guidance, enabling me to become a better researcher.

I value the opportunity given by the University of Nottingham to pursue a PhD and carry out this study. I would like to give a special thank you to the members of staff at the School of Economics, University of Nottingham for the valuable criticism and pleasant academic environment to conduct this research. To my colleagues and office mates (in Room B51), thank you for the encouragement and constructive feedback into this research. To the participants at various academic conferences and Centre for Research in Economic Development and International Trade (CREDIT) seminars, thank you for the helpful comments.

I acknowledge the support of the Department of Census and Statistics by providing the data used in this study. A special thank you to Mr. Migara Fernando who was kind enough to help me obtain the data without difficulty. I express my gratitude to Professor O. G. Dayaratna Banda (University of Peradeniya), Mr. J. Shivakumar (World Bank), Mr. Wasantha Dewage (Central Bank, Sri Lanka) and Miss Kalyani Rajapaksa for their useful insights on the Sri Lankan economy.

My sincere gratitude goes to my father, G. T. Bandara, my mother, Mangala Bandara and brother, Nivantha Bandara for their boundless love and support. A special thank you to my uncle and aunt - Rajasinghe and Rose Mudunkotuwe for being my family away from home throughout all these years. I would also like to thank my former teacher, Mrs. Varuni Jayasekara, who encouraged me to pursue this PhD. My friends, Nishan Fernando and Ashira Perera, thank you for always taking the time to provide the confidence and emotional support I needed to persevere.

I am truly blessed to have all of you in my life and thank you from the bottom of my heart.

Table of Contents

Abstract ..... **i**

Acknowledgements ..... **ii**

Chapter 1: Introduction of the Thesis **6**

1 Motivation **6**

2 Background of the Sri Lankan economy **8**

3 Data sets **12**

4 Thesis summary **14**

Chapter 2: Rates of Return to Education in Sri Lanka **17**

1 Introduction **17**

2 Literature review **19**

2.1 Returns to education: Theory and macroeconomic evidence 19

2.2 Microeconomic estimates of the returns to education 23

2.3 Returns to education in Sri Lanka 27

3 Data and methods **31**

3.1 Data 31

3.2 Ordinary Least Squares 35

3.3 Instrumental Variables 39

4 Results and discussion **43**

4.1 IV (in comparison to OLS) 43

4.2 Returns to education by level (OLS) 55

4.3 Production function at the household level 62

5 Conclusion **65**

6 References **67**

Appendix 1: First stage IV regression results **73**

Appendix 2: Other instruments **75**

Appendix 3: Variables used in the earnings and production functions **77**

Chapter 3: Decomposing the urban-rural welfare gap in Sri Lanka 80
1 Introduction 80
2 Background of the economy 82
3 Literature review 86
3.1 Theory and findings for other countries 86
3.2 Methods used in the literature 88
4 Data 91
4.1 Data and sample 91
4.2 Descriptive statistics 94
5 Determinants of urban and rural per capita expenditure 99
5.1 Method 99
5.2 Model specifications 100
5.3 Results 101
6 Factors contributing to the urban-rural gap in expenditure 113
6.1 Method 113
6.2 Results 115
7 Conclusion 124
8 References 127
Appendix 1: Data on the excluded districts 131
Appendix 2 133

Chapter 4: Sectoral choice and sample selectivity 149
1 Introduction 149
2 Literature review 152
2.1 Wage premium in the public sector 152
2.2 The role of the informal sector 155
2.3 Differences across men and women 158
3 Data and descriptive statistics 159
4 Methods 166
4.1 Sample selection 166
4.2 Endogenous explanatory variable(s) 169
5 Results 172
5.1 Determinants of sector of employment 173

5.2 Earnings across sectors 180
5.3 Predicted earnings 186
6 Conclusion 189
7 References 192
Appendix 1 197
Appendix 2: Blinder-Oaxaca decomposition 209

Chapter 5: Conclusion, summary of findings and scope for future research 215
1 Summary of findings and contributions to literature 215
1.1 Rates of return to education in Sri Lanka 215
1.2 Decomposing the urban-rural welfare gap in Sri Lanka 216
1.3 Sectoral choice and sample selectivity 218
2 Policy implications, limitations and scope for future research 220

Chapter 1: Introduction of the Thesis

1 Motivation

By the provision of skills and knowledge, human capital is seen to boost productivity levels in both, developed as well as developing countries. Education plays a vital role in providing access to better jobs and reducing income poverty. The returns on investment in education have been studied extensively in the literature (Mincer, 1974; Psacharopoulos, 1994; Psacharopoulos and Patrinos, 2004; Kingdon et al., 2008). These studies look at different countries, all confirming the positive link between education and earnings; educated individuals are more productive and are able to earn higher wages than their less-educated counterparts.

Arrow (1973) states that this link is valid only if the observed differences in earnings across groups of individuals with different levels of education reflect true differences in productivity, rather than differences in unobservable characteristics such as ability or family background. If such characteristics are linked to education and earnings, then ignoring them could lead to biased estimates of the returns to education. There is a vast literature on the issue of endogeneity in measuring education. Studies use various instruments to obtain a more accurate measure of education that account for unobservables. Several studies have suggested that family background has an effect on an individual's level of education – for example, children of more educated parents are likely to attain higher education (Blackburn and Neumark, 1993, 1995; Parker and Van Praag, 2006); Pencavel (1998) discussed that individuals are likely to get married to people with similar levels of education, behavioural traits and/or common preferences, thus spouse's education may be linked to own education.

The literature on the rates of return to education further explores the returns across the education distribution – that is, the returns to primary, secondary and tertiary education. It used to be argued that education and earnings share a concave relationship (Psacharopoulos, 1994; Psacharopoulos and Patrinos, 2004) – primary education has higher returns than tertiary education partly due to the latter being more

expensive. However, more recent evidence (for example, Colclough et al., 2010) state that the relationship is convex – returns to education increase at every level.

In many developing countries, education is more widely available in urban areas relative to rural areas (Knight and Song, 1999). The educational system in rural areas could be demand-constrained, due to higher opportunity costs of schooling and low benefits. On the other hand, education in rural areas could be supply-constrained due to an "urban bias" in policy making (Lipton, 1977) – preference in access to better and subsidized education services is given to people in urban areas. This will lead to spatial differences in poverty between urban and rural areas which could eventually slow down the development process in developing countries. Human capital theory states that education is crucial for economic welfare of a household, as it has a strong impact on labour productivity and success, employment opportunities and earnings of individuals in the household (Schultz, 1961; Becker, 1964). Nguyen et al. (2007), Hnatkovska and Lahiri (2013), and Thu Le and Booth (2014) identified education as a key factor that contributes towards differences in wages and welfare (as measured by household expenditure) across urban and rural areas.

Differences in human capital endowments have been linked to earnings differentials; earnings are likely to be higher for individuals with higher levels of education. This has implications on labour market outcomes, especially in developing countries where the labour market is segmented. Four segments have been identified for developing countries, namely, public, rural, private formal and informal. The segments vary by nature of the job contracts, seasonality of production, uncertainty of demand and the structure of earnings. The formal sector is regulated and subject to explicit work contracts between the employees and employers. On the other hand, the informal sector is not entirely subject to such contracts and government regulations – the latter sector constitutes individuals and small-scale enterprises that employ a relatively smaller number of employees. The informal sector has been defined by the International Labour Organization as institutions that are (1) not registered with the Department of Inland Revenue or included in the Employment Provident Fund, (2) have no records of formal accounts, and/or (3) have less than 10 regular employees.

The importance of the informal sector in a developing country has been identified in several studies. Bennell (1996) states that aggregating the rural, formal and informal sectors could lead to an imprecise representation of the labour market as the rural and informal sectors are likely to have different types of individuals working in them – for example, they may have lower levels of education compared to formal and public sector workers. Furthermore, individuals in self-employment are less likely to be linked to individual characteristics such as education or to a pay scale (Vijverberg, 1995). Thus, human capital endowments may help individuals acquire jobs in high-paying sectors of employment. Given that the labour market is segmented in developing countries, policies relating to labour allocation and income distribution need to take this into account. Research on this topic is useful to uncover the employment choices and earnings determination across various sectors of the labour market.

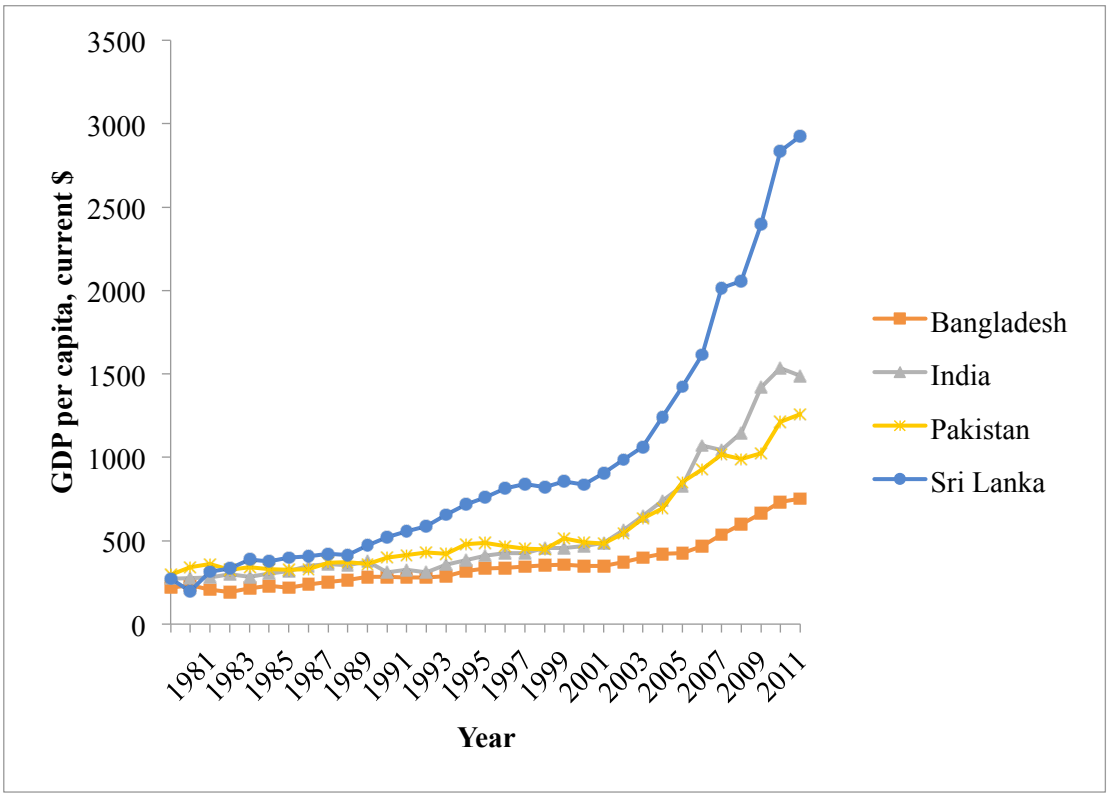
2 Background of the Sri Lankan economy

Before identifying the research questions, this section discusses some of the statistics, background information about the Sri Lankan economy, and where it stands relative to other comparable countries.

Sri Lanka has shown rapid growth over the last two decades. Figure 1 shows the country's growth, measured by GDP per capita, in comparison to three other South Asian countries – Bangladesh, India and Pakistan. Sri Lanka has had the highest GDP per capita since the 1990s. Considering the growth rate during the period 2005 to 2010, Sri Lanka has reported similar growth rates as India, at a rate of 10.4 per cent which is higher than Pakistan (3.6 per cent) and Bangladesh (8.8 per cent)[1].

[1] The growth rate over the 5 year-period has been calculated as follows: $\frac{\ln(y_{2010})-\ln(y_{2005})}{5}$ where "*ln* (*y*)" represents the natural logarithm of real GDP per capita

Figure 1: Real GDP per capita (current $)

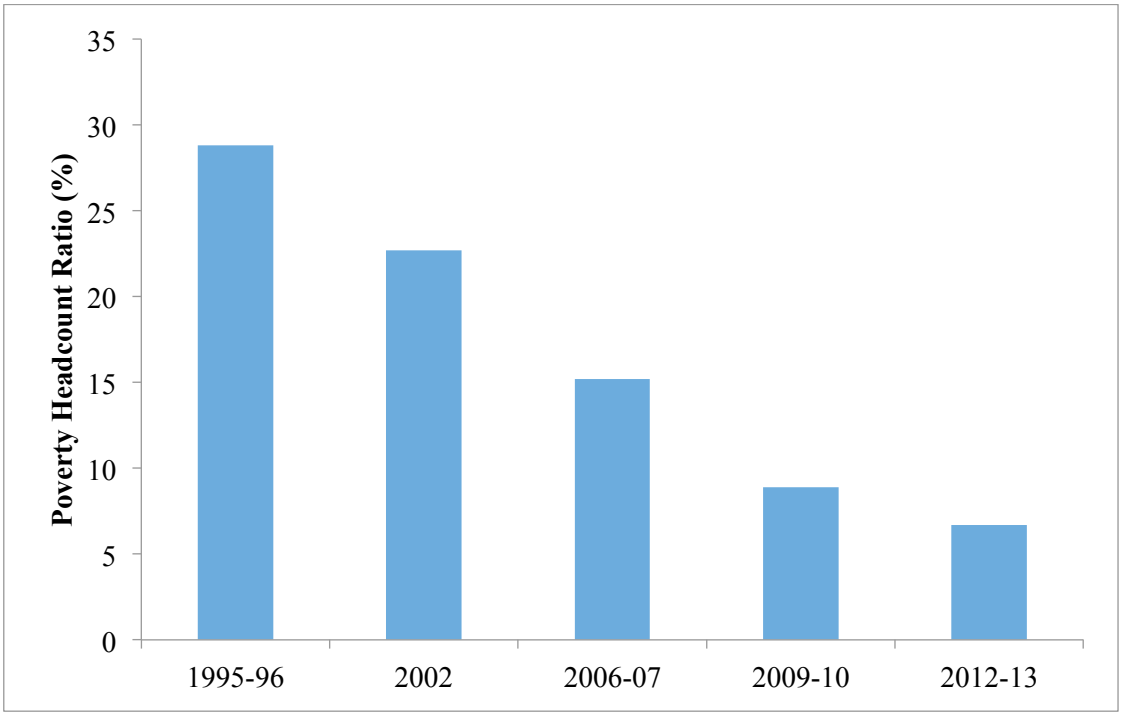


Source: The World Bank

As the country experienced growth, its poverty rates have been declining. Figure 2 shows the poverty trends in Sri Lanka, as measured by the poverty headcount ratio[2]. From 1995/1996 to 2012/2013, there has been an overall decline in this ratio (as shown in the graph below). Between 2002 and 2009/10, the poverty headcount ratio for the country has dropped drastically from 22.7 per cent to 8.9 per cent. The fall in poverty was primarily observed in rural areas – the poverty headcount ratio dropped from 24.7 to 9.4 per cent between 2002 and 2009/10 whereas the urban poverty headcount ratio fell from 7.9 to 5.3 per cent.

Figure 2: Poverty Headcount Ratio in Sri Lanka over the past 2 decades

[2] The poverty headcount ratio was measured at the national poverty line of 1,423 rupees (LKR) in



Source: Department of Census and Statistics

Sri Lanka's exports comprised of 20 per cent of total GDP in 2010, which is similar to India (22 per cent) and higher than Pakistan (13 per cent) and Bangladesh (16 per cent). The percentage of individuals employed in primary sectors (agriculture) was 32 per cent in 2010, with 26 per cent in the secondary sectors (industry) and 42 per cent in tertiary sectors (services). Compared to Bangladesh, India and Pakistan, the composition of Sri Lanka's work force is more service-oriented rather than agriculture-oriented. Across men and women, the sectoral composition is similar in the context of Sri Lanka; whereas in India, a relatively larger proportion of the female work force is employed in agriculture compared to men, and a smaller proportion in services. Other important differences have been noted across countries on female employment and wages. In 2010, over 33 per cent of the Sri Lankan labour force comprised of women. This is slightly lower than Bangladesh (40 per cent), but higher than Pakistan (22 per cent) and India (25 per cent). Riboud et al. (2007) observed gender-related wage differentials of 50 per cent in Sri Lanka and India, and almost 300 per cent in Bangladesh and Pakistan.

Sri Lanka has been practicing a free education policy since 1938, from the level of kindergarten until the tertiary level (undergraduate studies). In India, free and compulsory education for children between the ages of 6 and 14 was implemented in

2009, whereas in Bangladesh, compulsory education was introduced in 1972. The percentage of the population over the age of 15 that has no education is lower in Sri Lanka compared to other South Asian countries such as Bangladesh, Pakistan and India (4.4 per cent, compared to over 30 per cent in other countries for the year 2010). The proportion of the Sri Lankan population to have completed tertiary education in 2010 was almost 9 per cent, compared to 2-5 per cent in other countries. The country outperformed in terms of the youth literacy rate (individuals between the ages of 15 and 24) – 98 per cent, compared to 86 per cent in India, 71 per cent in Pakistan and 81 per cent in Bangladesh.

In 2010, 8.6 per cent of total government expenditure was spent on education. This is relatively low compared to other South Asian countries such as India (11.7 per cent), Pakistan (11.9 per cent) and Bangladesh (14 per cent). Out of the total expenditure on education, 53 per cent was spent on secondary education, 25 per cent on primary and 16 per cent on tertiary – compared to Sri Lanka, India spends relatively more on tertiary education (36 per cent of total education expenditure) and less on secondary education (37 per cent). In 2010, Sri Lanka's government expenditure per primary student as a proportion of GDP per capita was 5.9 per cent (7.2 and 7.9 per cent in India and Bangladesh respectively), per secondary student was 8.6 per cent (13.6 and 10.6 per cent in India and Bangladesh respectively) and per tertiary student was 25.5 per cent (68.7 per cent in India). This suggests that the country has a more educated population compared to some South Asian countries, but the expenditure on education has been relatively low.

Additionally, the country had seen the end of a 30-year civil war during this period. As a result of this, it was able to use resources which had previously been prioritized for the war to implement policies concerning development of the economy. By 2010, the access to drinking water, electricity and other facilities improved for rural households. Further, rural development and infrastructure projects commenced across all areas (World Bank, 2009).The transportation sector (particularly covering rural areas) was largely seen in the road development sector, with the expansion of transport services. Improvements were made in the road network to ease passenger and goods transportation. Due to the damage caused to roads because to the Tsunami

in 2004, a programme to revitalize roads in rural areas commenced afterwards. This enabled easy movement for individuals between towns and villages.

As a developing country that has seen progress in recent years, the case of Sri Lanka offers us the opportunity to expand the literature on human capital and labour market outcomes. With this in mind, this research aims to answer the following questions:

1. What is the effect of education on earnings and productivity? How does the effect vary across different levels of education (primary, secondary or tertiary)?

2. How has the urban-rural welfare gap changed over recent years, and what has caused this change?

3. What factors determine an individual's choice of sector? Do earnings differ across sectors, for both, men and women?

3 Data sets

Three data sets are used to address the above issues – the Sri Lankan Household Income and Expenditure Surveys (HIES) for 2002 and 2009/10, and the Quarterly Labour Force Survey (QLFS) for 2013. The HIES for 2002 was conducted during the period from January 2002 through December 2002, while the 2009/10 survey was conducted during July 2009 through June 2010. The QLFS was conducted for the period from January 2013 through December 2013.

There are nine provinces and 25 districts in the country. Due to the war and resettlement activities in the Northern and Eastern provinces, the HIE survey for 2002 excluded the districts in these two provinces. For the survey held in 2009/10, data collection commenced in some of these areas – however it did not cover the entire 12 months. Therefore, we exclude the Northern and Eastern provinces in the analysis to have a comparable sample over the years. The QLFS held in 2013 included all 25 districts.

All three surveys use a two-stage stratified sampling procedure. The urban, rural and estate[3] sectors in each district are used as the selection domains; thus the district is the main domain used for stratification. For the HIE surveys (2002 and 2009/10), the sampling frame was a list of housing units prepared for the Census of Population and Housing (CPH) in 2001. The 2009/10 survey was the last to use this sampling frame. Therefore, the QLFS 2013 used a new sampling frame prepared by the CPH in 2011 based on a newly prepared set of census blocks. In the first stage, the sampling frame (that is, a collection of census blocks) is used for the selection of primary sampling units (PSUs). Within each selection domain, 2,500 PSUs are selected – the selection of the PSUs is done by assigning a probability of selection to each census block that is proportionate to the number of housing units in each census block within the selection domains. The selected PSUs are kept up-to-date to include new housing units and exclude demolished/vacated units prior to the survey. In the second stage, 10 secondary sampling units (that is, housing units) are selected from the 2,500 PSUs selected in the first stage – each housing unit in the PSU is given an equal probability of being selected for the survey. At the end of the selection process, a sample of 25,000 housing units is obtained to represent the whole country with different probabilities given based on the different sample sizes of the selection domains.

To address the first question in the thesis mentioned above, the HIES for 2009/10 is used. To estimate the earnings functions, all individuals in the workforce (wage and self-employed) are included - this comprises of a sample of 23,414 individuals. The sample for the production functions comprises of 3,460 and 3,320 households involved in agricultural and non-agricultural self-employment, respectively. As the HIE survey gives detailed information on income, output and consumption of the self-employed, it serves the purpose of this analysis.

To estimate the urban-rural welfare gap over time as the second aim of this thesis, the HIES for 2002 and 2009/10 are used comprising of 16,920 and 17,182 households respectively. The HIE surveys include detailed information on expenditure (on both, food and non-food items), making the surveys useful for this analysis.

[3] The estate sector is defined as plantations with at least 20 acres of land with 10 or more resident labourers.

To address the final question on labour market outcomes, the Quarterly Labour Force Survey for 2013 is used. The data set comprises of 69,253 individuals, out of which 25,382 are in the labour force. This survey includes information on hours worked by those in the labour force – this information is not available in the HIES series and is useful in the analysis of labour force outcomes as we are able to estimate earnings per hour. It also includes information that enables us to isolate the informal sector, which is not available with the HIE surveys. Further, the 2013 survey is the first in its series to include information on the income of the self-employed. Therefore the analysis examines several sectors in isolation.

4 Thesis summary

To examine the first issue, we estimate the Mincerian earnings model where the rates of return to education by gender and employment type (agricultural or non-agricultural activities in the public or private sector) are analysed. This chapter contributes to existing literature on the rates of returns to education. To the best of our knowledge[4], there is no existing study that explores the endogeneity of education for the case of Sri Lanka. An Instrumental Variables (IV) approach is applied in order to address the issue of endogeneity in the specification of years of education. Two variables are used as instruments for education – spouse's level of education and a policy variable that measures the change in the minimum years of compulsory schooling implemented in Sri Lanka in 1997. Harmon and Walker (1995) used a similar policy instrument – they relied on the exogenous change in the education distribution in the U.K. due to an increase in the schooling leaving age from 14 to 15 in 1947, and a further increase to 16 in 1973. Several studies have used spouse's education as an instrument to account for endogeneity of individual's level of education (for example, Trostel et al., 2002). Following such studies, similar instruments are used in this analysis. This enables us to contribute to the literature on the rates of returns to education by identifying whether there is a need to estimate education as an endogenous variable. If there is such a need, it would suggest that the OLS estimates are biased.

[4] Based on a literature search in EconLit

The second part of the analysis employs an Ordinary Least Squares (OLS) estimation to examine the returns at different levels of education – primary, secondary and tertiary. Due to a lack of instruments, we are unable to carry out an IV estimation here. However, this analysis sheds light on the convexity or concavity of the relationship between education and earnings. Additionally, we are able to identify how the returns vary across men and women, and workers in different types of employment – perhaps women receive higher earnings than men at certain education levels. We update the findings of Riboud et al. (2007), by using a more recent data set (the Household Income and Expenditure Survey for 2009/10) and by disaggregating the sample by gender and type of employment.

The final part of the analysis employs a production function to examine the productivity differences across households involved in agricultural and non-agricultural self-employment. Consequently, we include land, labour and raw materials/capital as additional determinants of agricultural/non-agricultural output.

The next topic analyses the urban-rural welfare gap in Sri Lanka for the years 2002 and 2009/10 using the relevant Household Income and Expenditure Surveys. To the best of our knowledge[5], the welfare gap has not been formally analysed so far in the case of Sri Lanka; the recent developments suggest that it is necessary to do so. This is the first contribution to the literature. Secondly, the chapter adds to the existing literature through the use of a new method by Firpo et al. (2009) allowing us to examine the determinants of expenditure and decompose the urban-rural welfare gap across the entire distribution of expenditure. It is believed that the welfare gap varies across the expenditure distribution, thus an evaluation at the mean seems inadequate. Therefore an unconditional quantile regression is employed; the method enables us to identify the determinants of urban/rural household expenditure at different quantiles. After identifying the determinants across the distribution for urban and rural households, we are able to isolate the differences in characteristics (for example, education) and returns to characteristics between urban and rural households that contribute towards the overall urban-rural expenditure/welfare gap. A variant of the threefold Blinder-Oaxaca decomposition is used for the analysis of the welfare gap.

[5] Based on a literature search in EconLit

To address the final question, the labour market outcomes in Sri Lanka are analysed using the Quarterly Labour Force Survey for 2013. Firstly, the individual determinants of labour force participation and sector choice are estimated. The second part of the analysis estimates the earnings functions for individuals working in different sectors of employment. Five sectors are identified – agriculture, non-agriculture public, private formal and informal wage earners, and self-employed. Additionally, we are able to control for endogeneity of education using a Control Function (CF) approach following the works of Garen (1984), Wooldridge (2005), Söderbom et al. (2006) and Kuépié et al. (2009). The CF regresses education on all exogenous variables; residuals are obtained and are then included in the multinomial logit and earnings functions. Hence, the first part of the analysis looks at the determinants of employment sector for men and women. Following Heckman (1979) and Lee (1983), we address the potential issue of sample selectivity. As we are unable to observe wages for individuals who are not in the labour force, we estimate earnings for a selected, non-random sample. In order to ensure that this does not lead to biased estimates of the earnings functions, we deal with this potential issue of selectivity. Multinomial logit estimation is used, which includes individuals who are not in the labour force. Thus, we are able to estimate the earnings functions for individuals in each sector, after accounting for endogeneity of education and controlling for any form of selectivity bias.

This chapter contributes to the existing literature by addressing two forms of possible bias simultaneously – endogeneity of education and selectivity biases in the earnings functions for individuals in different sectors of employment. Further, the chapter adds to the existing literature on labour market outcomes in Sri Lanka.

The rest of the thesis is organized as follows. Chapter 2 addresses the first research question by estimating the earnings functions to examine how the rates of returns to education vary across sub-groups. Chapter 3 explores the urban-rural welfare gap in Sri Lanka between 2002 and 2009/10 to identify the factors that contribute towards the gap in both periods. Chapter 4 analyses the determinants of employment sector choice, and the implications it may have on earnings. Finally, chapter 5 concludes by summarizing the key findings, contributions and policy implications.

Chapter 2: Rates of Return to Education in Sri Lanka

1 Introduction

This chapter examines the returns to education across the sexes and by different types of employment. It then investigates how the returns vary at different levels of education (primary, secondary and tertiary) in Sri Lanka during 2009/10. Additionally, the chapter deals with the issue of endogeneity of education when estimating earnings. This has not been explored in the context of Sri Lanka[6]. 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.

From developed to less-developed countries, it has been shown 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.

[6] To the best of our knowledge, based on a literature search in EconLit

Building on a large literature on the returns to education, this chapter 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. 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 potential 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

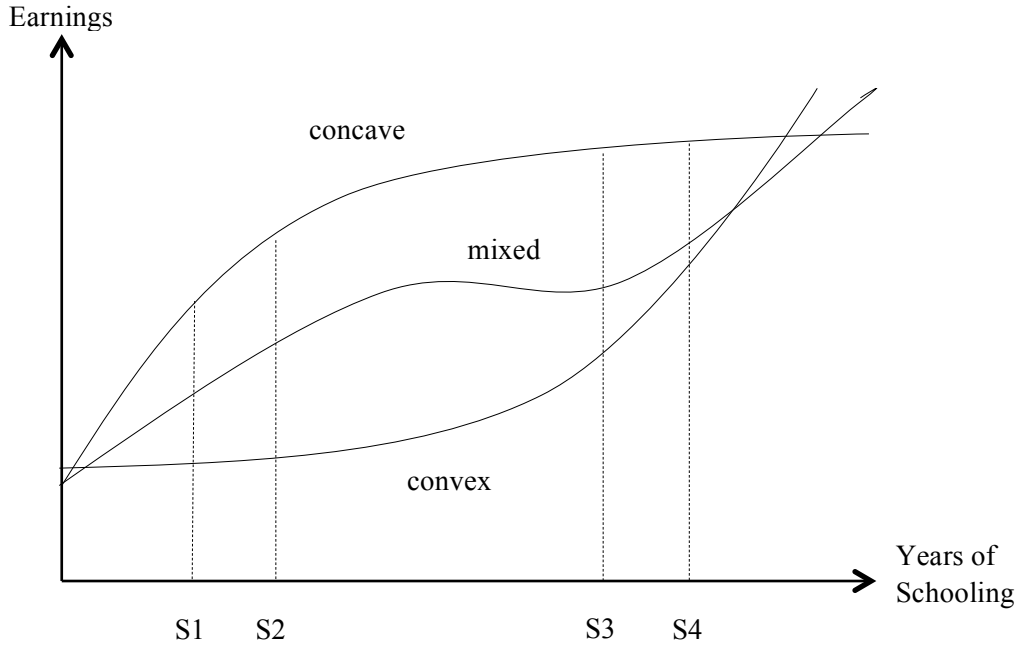
such, social benefits are realised. Colclough et al. (2010) state that the social rates of return to education are lower than the private rates of return to education possibly due to the addition of "publicly financed costs" to the private returns. The entire economic impact of education may not be captured 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, low crime rates.

This chapter focusses 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[7].

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.

[7] Henceforth, this chapter refers to the Mincerian returns to education, unless stated otherwise

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. 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[8]. 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

[8] 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.

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.

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[9], 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[10]. 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 which 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[11]. Figure 2 (b) looks at education expenditure per student as a share of GDP per capita[12]

[9] Based on a search in EconLit

[10] 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.

[11] Similar patterns were seen for the previous years – 2008, 2009 and 2011

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

- 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.

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

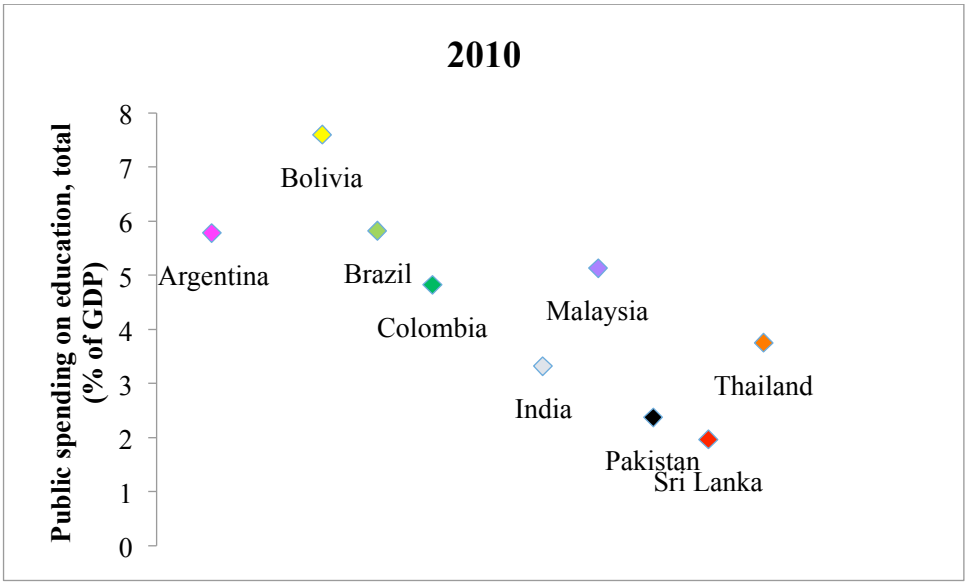
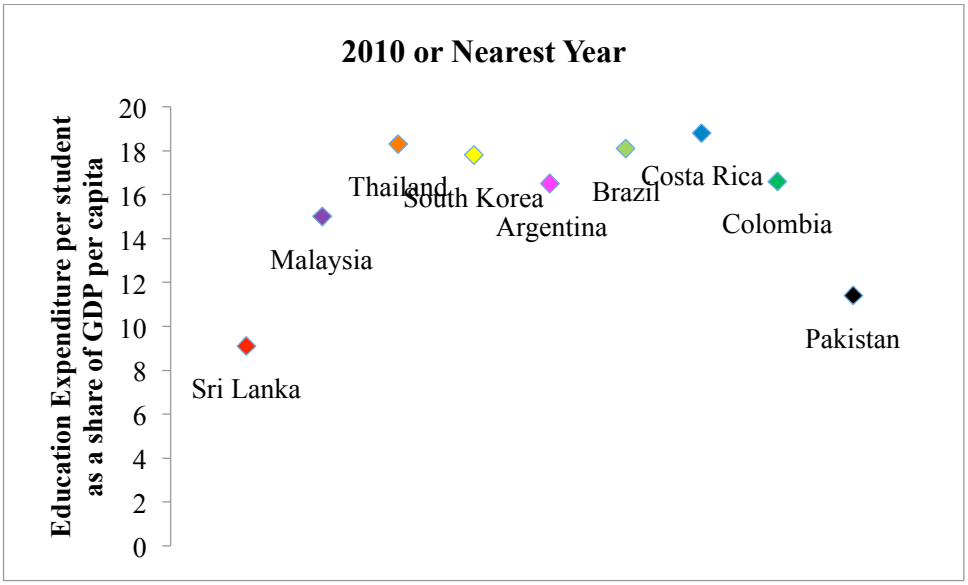


***Figure 2(b)**: Education Expenditure per student (as a share of GDP per capita)*



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). 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[13]. 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

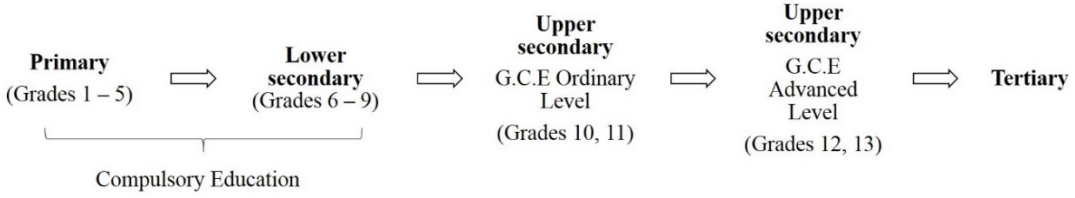
[13] The returns to education are lower when all districts are included – for the full sample, as well as for the sub-samples by gender and type of employment.

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. 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[14].

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

[14] 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.

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[15]. 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.

[15] The informal sector will be examined in chapter 4

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 chapter 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[16] 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[17]. 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.[18] Some individuals (2012 observations) 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:

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

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

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

[18] 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.

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[19] 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[20]). 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,

[19] 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

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

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 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[21]. 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.

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

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

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[22] and labour. Unfortunately, there is no detailed breakdown of the cost of inputs, but the sum of capital and raw materials (such as seeds and fertilizer)[23].

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.

[22] 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

[23] 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

3.3 Instrumental Variables

As discussed in Section 2.2, a potential issue with OLS is that "S", representing the aggregate years of schooling[24], 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[25]:

$$log\ _Y_i = \beta_0 + \beta_1 \widehat{S_\iota} + \beta_2 W_i + \varepsilon_i$$

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:

- 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.

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

[24] 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.

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

- 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 chapter, 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)[26]. 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.

[26] 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[27].

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

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) which includes accounting for the occupation and industry of the individual.

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

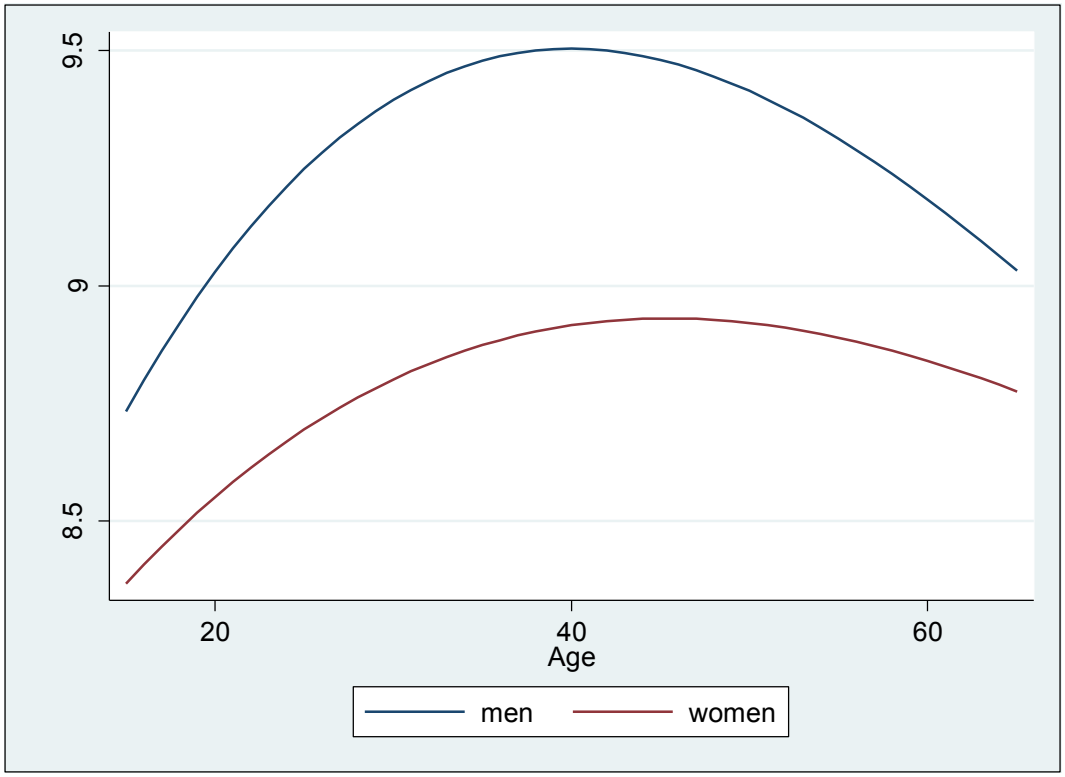
For the purpose of comparing IV with OLS, we focus on the Mincerian returns (given by specification (1) in the tables). For men, the IV estimate for the return to education is 12.6 per cent, higher than the OLS estimate of 8.6 per cent. Similarly for women, the IV estimate is 14.8 per cent while the 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.

Figure 4 shows an inverse U-shaped relationship between the age variables and (predicted) earnings[28] 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,

[28] 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

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



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.

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. We will explore this further in chapter 4.

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

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

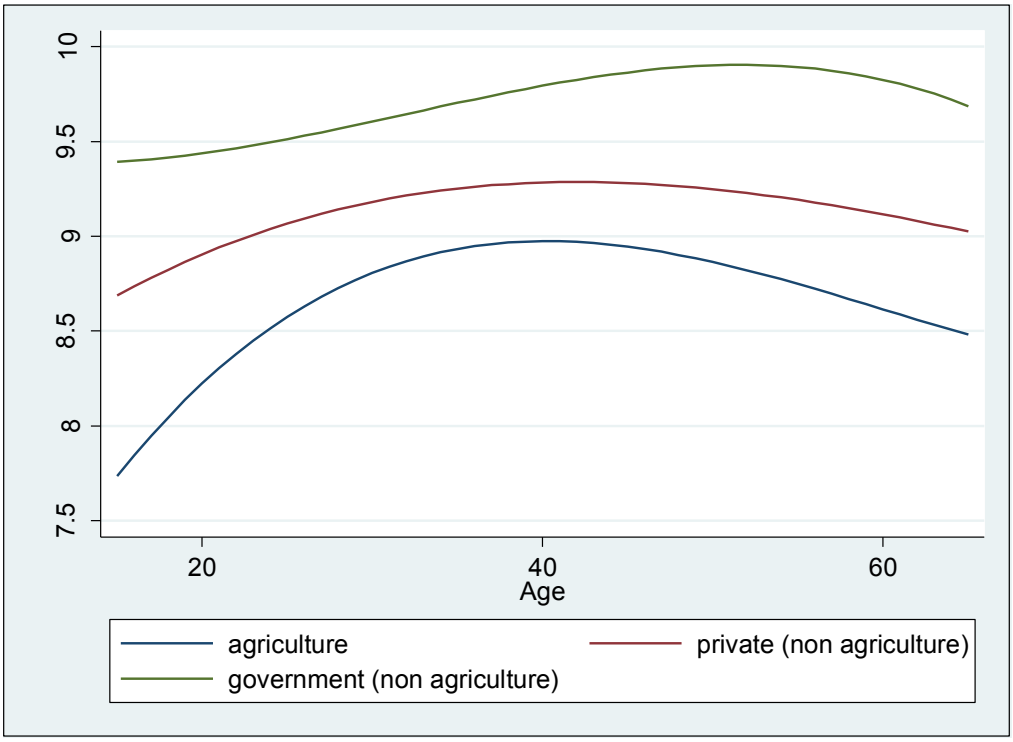


Figure 5 presents the age-earnings profiles for the three different sub-divisions using the IV estimates. Across the age distribution, the predicted earnings are highest for government sector workers and lowest for agricultural workers. For agricultural and private sector workers, the peak of earnings 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.

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

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.

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. We will discuss this further in chapter 4.

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[29]. 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.

[29] 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.

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).

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 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.

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.

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

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[30] - that is, they have more opportunities in

[30] 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.

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

5 Conclusion

This chapter 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 chapter does not account for the non-random way in which individuals choose into various sectors of employment. Chapter 4 deals with this potential bias of sample selectivity in greater detail.

6 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”, *the 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", *the 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*** (9.72)	8.074*** (10.64)	2.125 (1.54)	4.802*** (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 (-0.62)	0.179*** (3.01)	0.054 (1.07)	0.669*** (3.10)	0.286* (1.58)
Age^2	0.001 (0.56)	-0.005*** (-4.17)	-0.002** (-2.17)	-0.016*** (-2.98)	-0.007* (-1.72)
Age^3	$-1.07x10^{-5}$ (-0.71)	$4.07x10^{-5}$*** (4.35)	$1.76x10^{-5}$** (2.24)	$1.13x10^{-4}$*** (2.72)	$5.58x10^{-5}$* (1.66)
Female	0.246** (2.08)	-0.026 (-0.47)	-0.104** (-2.02)	0.757*** (8.24)	-0.132* (-1.57)
Urban	0.877*** (4.24)	0.934*** (17.49)	0.394*** (8.12)	0.594*** (6.31)	-0.020 (-0.25)
Married	-4.016*** (-20.99)	-4.666*** (-42.59)	-3.379*** (-31.78)	-7.076*** (-26.16)	-3.828*** (-16.34)
Sinhalese	1.080*** (5.93)	1.491*** (24.73)	1.138*** (20.64)	2.765*** (16.20)	1.110*** (7.51)
Occupations:					
Managers			2.352*** (16.45)		3.709*** (12.58)
Professionals			4.303*** (25.34)		4.187*** (21.02)
Technicians			2.329*** (17.49)		2.915*** (14.66)
Clerks			3.688*** (25.76)		2.502*** (12.90)
Services			1.776*** (12.58)		1.271*** (6.42)
Craft			0.924*** (7.23)		0.853*** (3.14)
Operators			1.358*** (10.30)		0.896*** (3.89)
Elementary			-0.121 (-1.05)		-0.084 (-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)	-0.119*	-1.86		
Quarter 2	0.140**	2.18		
Quarter 3	0.084	1.32		
Quarter 4				
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	23,414	0.10	0	1	0.31

	professional					
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 of Sinhalese employed persons	3,460	0.95	0	1	0.22

(proportion)						
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)	average years of upper secondary schooling acquired by employed persons	3,460	1.14	0	4	1.27
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

Chapter 3: Decomposing the urban-rural welfare gap in Sri Lanka

1 Introduction

This chapter explores the urban-rural welfare gap in Sri Lanka between 2002 and 2009/10, a period of high growth and falling poverty rates. It aims to answer three questions – (1) what are the determinants of urban/rural household welfare, (2) how has the urban-rural welfare gap changed over time, and (3) what factors contribute towards the widening or narrowing of the gap. Welfare is measured using household expenditure per capita. In order to explore urban-rural welfare differences across rich and poor households, we examine the entire expenditure distribution.

The literature views economic policy as being subject to an "urban bias". Knight and Song (1999) define the urban bias as the government being more concerned about urban development rather than rural development due to the political power of urban dwellers. Lipton (1977) who popularized this concept noted that spatial differences in poverty across urban and rural areas will slow down the growth process in developing countries. Sri Lanka has experienced high economic growth, falling poverty rates and is transitioning from being an agricultural- to a service-oriented economy over the last decade. The poverty headcount ratio[31] fell rapidly from 22.7 per cent to 8.9 per cent between 2002 and 2010, primarily due to the sharp fall in rural poverty (Department of Census and Statistics). In early 2009, Sri Lanka saw the end of a 30-year war that had restricted development of the Northern and Eastern provinces. The country began to prioritize rural development and new infrastructure projects commenced across all areas (World Bank, 2009). In developing countries, the urban-rural welfare gap is considered a vital element of inequality (Nguyen et al, 2007; Lahiri, 2013; Thu Le and Booth, 2014). Understanding the magnitude of the welfare gap as measured by

[31] The poverty headcount ratio was measured at the national poverty line of 1,423 rupees (LKR) in 2002, and 3,028 rupees in 2010 (base year is 2002). The poverty headcount ratio at $1.90 a day (2011 PPP) was 8.25 per cent in 2002 falling to 2.41 per cent in 2010. The Northern and Eastern districts have not been included due to availability of limited data prior to 2009/2010

household expenditure per capita and identifying the key factors contributing to this gap is the goal of this chapter.

The chapter contributes to the existing literature in two ways. First, the urban-rural welfare gap has not been formally decomposed for the case of Sri Lanka[32]. This chapter uses data from the 2002 and 2009/10 Household Income and Expenditure Surveys covering the significant period of Sri Lanka's growth and transition. The second contribution is through the use of a new method of unconditional quantile regression (Firpo et al., 2009) applied to examine the determinants of per capita expenditure for urban and rural households across the expenditure distribution. Using this method enables an isolation of the factors that contribute to the urban-rural expenditure gap across the expenditure distribution. To do this, the Blinder-Oaxaca decomposition is applied directly to the estimation results from the unconditional quantile regression. The benefit of this method is that it does not require the estimation of several simulations; which are necessary with the alternative method of conditional quantile decomposition. The estimation of the Blinder-Oaxaca decomposition allows us to answer two of the fundamental questions in this chapter – (1) how does the welfare gap vary across the expenditure distribution, and over time, and (2) what factors contribute towards this gap. The Blinder-Oaxaca decomposition isolates the urban-rural differences in characteristics/endowments (for example, differences in education levels across the two groups) from the urban-rural differences in returns to such characteristics (differences in the returns to education across urban and rural areas).

The findings obtained in this chapter suggest that individual characteristics such as education, employment in the services sector, the presence of children in the household and receiving remittance income have a positive association with per capita household expenditure, whereas household size and employment in the agriculture sector have a negative association with per capita expenditure. This is true for households in both urban and rural areas. We then estimate the urban-rural welfare gap. At a given point in time, the gap increases from the bottom to the top of the expenditure distribution indicating that the differences in per capita expenditure

[32] This was revealed from a literature search in EconLit

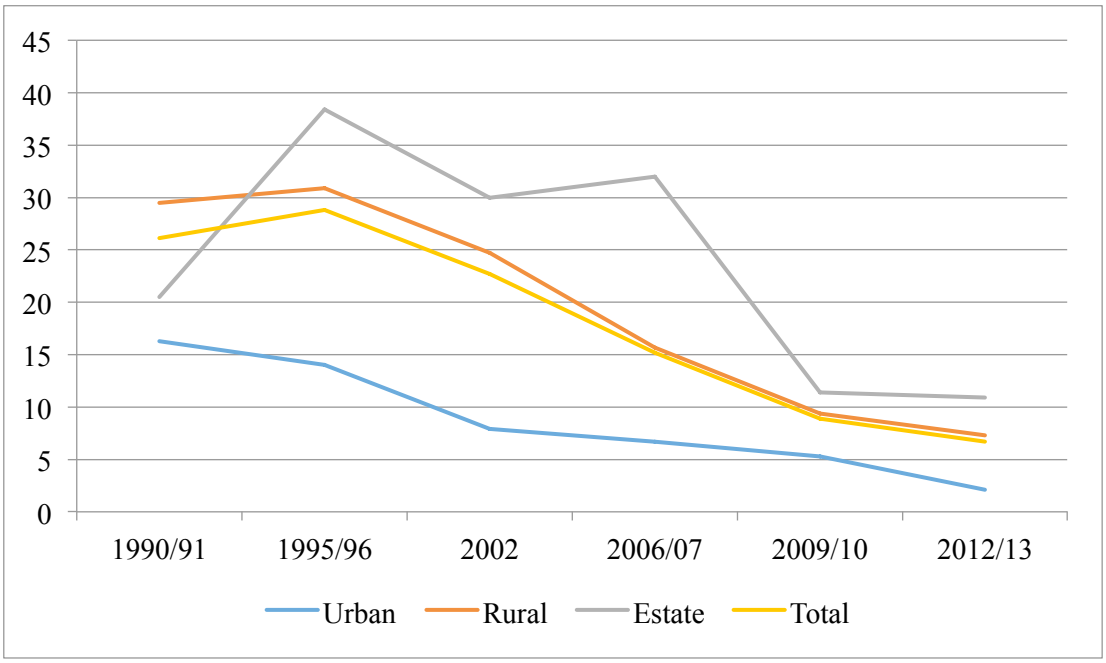
between urban and rural households are greater for richer households. A number of factors contribute to the urban-rural welfare gap, including urban-rural differences in the levels of education, regions, industrial structure, and the relevant returns from these factors. Between 2002 and 2009/10, the urban-rural welfare gap reduced greatly; this was primarily due to lower urban-rural differences in returns to education, ethnicity and other individual characteristics. Overall, an adjustment of the total endowments/characteristics of rural households to those of urban households reduces the urban-rural expenditure gap by approximately 43 per cent in 2002 and 50 per cent in 2009/10 (at the median).

The rest of the chapter is organized as follows. Section 2 summarizes the recent growth and development in Sri Lanka. Section 3 reviews some of the literature related to the urban-rural gap. Section 4 examines the data and variables that will be used in this study, followed by Sections 5 and 6 which employ the unconditional quantile regression technique introduced by Firpo et al. (2009) and apply the regression results to the Blinder-Oaxaca decomposition. Section 7 concludes along with policy implications and scope for future research.

2 Background of the economy

The Sri Lankan economy has undergone several changes in recent years. The aim of this section is to identify why an analysis of the urban-rural welfare gap during a period of development is worth examining. We explore the changes in poverty rates, urban/rural living conditions, regional differences and the changes in the industrial structure between 2002 and 2009/10. By doing so, we are able to identify some of the important changes that took place in this economy over time. In Sri Lanka, poverty as measured by the headcount ratio has dropped over time. The substantial fall in rural sector poverty is the leading contributor (82 per cent) to the drop in poverty at the national level between 2002 and 2009/10. Figure 1 shows the changes in poverty over time for each of the three sectors. Between 2002 and 2009/10, the rural poverty headcount ratio fell dramatically from 24.7 to 9.4 per cent, urban poverty headcount ratio fell from 7.9 to 5.3 per cent and estate poverty fell from 30 to 11.4 per cent.

Figure 1: Poverty Headcount Ratio (percentage)



Source: Department of Census and Statistics (author's computation)

The decrease in poverty has been linked to the rural development and infrastructure improvements across the country. Sri Lanka has had several projects targeted at the improvement in access to basic facilities for rural households. The Rural Hydro Electricity and water projects are some of the major projects; for example, the rural water project aided over 400,000 rural households in obtaining access to clean drinking water. Since the Tsunami affected many coastal areas in 2004, several development projects were undertaken to transform rural areas. Another project, the Samurdhi program, was established to encourage participation of the poor, expanding opportunities for self- and wage employment at rural levels.

Statistics from the Censuses of Population and Housing for 2001 and 2011 show vast improvements in the access to various facilities such as clean water, electricity and the major sources of cooking fuel. While this is especially true for rural and estate areas, it is also the case that urban areas saw improvements. In 2001, 85.3 per cent of urban households had electricity as the main source of lighting and this rose to 96 per cent in 2011. In the rural sector, 62 per cent of households had electricity as the major source of lighting in 2001. By 2011, 85 per cent of the rural households had electricity. The main source of drinking water used by urban households is piped-born water (77.8 per cent of the households in 2001 and remained quite stagnant even in 2011). For the

rural sector, the main source of drinking water comes from wells; in 2001, 58 per cent of rural households drank from protected wells and 11.5 per cent drank from unprotected wells. By 2011, the corresponding figures were 54 per cent and 4.8 per cent. In the urban sector, households use gas as the main source of cooking fuel and this rose from 45.3 per cent to 53.7 per cent of the population using it between 2001 and 2011. For rural and estate households, the main source is firewood, but more households report using gas as the main form of fuel in 2011. As the gap in terms of access to major household facilities has narrowed in recent years, we observe in our analysis that the urban-rural welfare gap has also narrowed over time.

In the past few years, road and air transportation recorded significant growth. The development in the transportation sector (particularly covering rural areas) was largely seen in the road development, with the expansion of transport services (both, rail and road). Improvements were made in the road network to ease passenger and goods transportation. A programme to revitalize roads in rural areas that commenced in 2004 was able to rehabilitate over 840 kilometres of roads (Central Bank of Sri Lanka, 2012). This focus on rural infrastructure development became a priority after the Tsunami which damaged significant areas of the rural transport sector. This enabled easy movement for individuals between towns and villages.

The country however has seen persistent differences across regions; the Central, Sabaragamuwa and Uva provinces (that include a large proportion of plantations) in particular, still have high levels of poverty. UNDP (2012) argues focus on making improvements at the regional level is required to prevent inequality from limiting the development of the economy. Regional differences also explain the significant variances in urban growth rates, and the diverse levels of urbanization among districts. Table A2 (refer to Appendix 2) presents the urban population by district. Uduporuwa (2010) showed that the Western province is the core urbanized region with the higher number of urban centres and highest percentage of urbanization. Other provinces however have not achieved significant urban growth. Additionally, the country having recently seen the end of a 30-year war meant that development in the Northern and Eastern provinces was limited for a long period of time and skilled labour migrated to urban areas in the Western and Central provinces for better

employment opportunities. This variation in levels of urbanization across districts and provinces motivates the inclusion of district-level variables in the analysis.

Having discussed the development of rural areas and regional differences that persist in Sri Lanka, we now turn to the changes in the industry structure that have taken place in the period under investigation. As we will observe, several changes occurred in terms of sectoral contribution to GDP and GDP growth, strengthening the argument for examining welfare changes over time. Table 1 gives the sectoral contribution to GDP and the growth rate in each sector (agriculture, industry and services) during the period of analysis. The contribution of agriculture to real GDP has fallen over time, remained fairly stagnant for industry and risen for services. The growth rate in GDP for each sector between 2002 and 2010 suggests that agriculture has the lowest growth rate of 3.8 per cent, whereas industry and services have the highest (6.3 and 6.4 per cent, respectively). This establishes the fact that not only do services contribute largely to GDP at a given point in time (over 59 per cent), but the growth in services contributes to overall growth of the economy. The last column measures the sectoral share of growth between 2002 and 2009/10. Services had the largest contribution towards GDP growth (62 per cent), followed by industry (30 per cent) and agriculture (8 per cent).

Table 1: Sectoral contribution to real GDP

	2002		2010		Growth between 2002 and 2010 %*	Share of growth by sector %*
	% share of GDP	2002/01 growth %	% share of GDP	2009/10 growth %		
Agriculture and Fishing	14.3	2.0	11.9	7.0	3.8	8
Industry	28.0	3.6	28.7	8.4	6.3	30
Services	57.7	4.8	59.3	8.0	6.4	62
Total	100	4.0	100	8.0	6.0	100

* Compiled by Author

GDP growth between 2002 and 2010 in each sector is measured as $\frac{ln(y_{2010})-ln\,(y_{2002})}{8}$ where "y" is the real GDP

The share of GDP growth by sector is measured as: $\frac{change\ in\ real\ GDP\ (by\ sector)\ between\ 2002\ and\ 2010}{change\ in\ real\ GDP\ between\ 2002\ and\ 2010}$

Population growth over the period 2002 to 2010 is 1 per cent per annum

Source: Department of Census and Statistics (DCS) National Accounts

Despite the strategies that were implemented to boost economic growth and reduce poverty to a large extent, persisting differences were observed across regions and socio-economic groups of the population. The Western province (including Colombo) has better access to markets, improved infrastructure, a greater proportion of educated people, and is dominated by non-agricultural sectors compared to other provinces. The above findings suggest that the country has gone through several changes in recent years. This gives rise to the key issue to be examined in this chapter – analysing welfare differences across urban and rural areas, and how this has changed over time.

3 Literature review

3.1 Theory and findings for other countries

This chapter tries to answer the central question of what factors contribute to the urban-rural welfare gap in Sri Lanka. The motivation for doing so is both theoretical and methodological.

Theoretically, over the years, two main frameworks have been central to the study of the urban-rural divide: the view that economic policy is subject to an "urban bias", and the Lewis model of surplus rural labour. The concept of an urban bias was popularized by Lipton (1977) who noted that spatial differences in poverty between urban and rural areas slows down the development process in poor countries. Subsequently, the urban-rural gap in welfare and income has been of concern in development research. The theory of an urban bias stems from the notion that the government favours urban areas over rural areas because of the political power of urban dwellers (Knight and Song, 1999). Despite the urban dwellers being a small proportion of the total population in developing countries, their influence on government policy is argued to be disproportionate to their numbers. If the government permits a wage differential favouring urban employment, rural-urban migration will be in excess of the capacity of urban areas which in turn gives rise to urban unemployment.

The two-sector model by Lewis (1954) was originally based on the argument that developing countries have a surplus of unproductive labour in agriculture primarily in rural areas which can be shifted to the growing manufacturing sector in urban areas to promote industrialization and sustainable development. Improving urban areas in this manner was thought to be efficient, but could come at the cost of national equity. Over time, Lewis's thinking about the role of agriculture shifted towards an emphasis on the increase of agricultural productivity and demand for agricultural goods (Lewis, 1978). This view was shared by others such as Mellor (1976) and Meier (1989) who identified the importance of agriculture not only as surplus to support industrialization, but also to view it as an activity by itself that generates employment, growth and a more equal distribution of income.

In many developing countries, the urban-rural welfare gap accounts for an important element of inequality. Vietnam is a case in point and a useful comparator for Sri Lanka given the countries' structural similarities. In a study on the Vietnamese urban-rural gap in welfare between 1993 and 2006, Thu Le and Booth (2014) employed a quantile regression technique and found that Vietnam's economic reforms such as the achievement of macroeconomic stability and the transition from a centrally-planned economy to a market economy in 1986, enabled households in urban areas to reap the benefits of the reforms (via higher returns to education) more than households in rural areas. Vietnam, like Sri Lanka experienced exceptional growth, but a rising urban-rural gap in welfare. However, since 2002, the urban-rural gap started to decline due to the development and industrialisation of rural areas. During the latter period, the urban-rural gap fell during a period of high growth bringing the rural households closer to urban households in terms of welfare.

Understanding the underlying factors affecting the urban-rural welfare gap is central to this chapter. The growth of certain industries, education and other household characteristics have been identified in the literature as potential contributors to the gap in welfare between urban and rural households. Thu Le and Booth (2014) applied the Blinder-Oaxaca decomposition to the unconditional quantile regression which identified the crucial role played by remittances and the loosening of government controls allowing rural migrants to access urban facilities such as education, health

insurance and owning a house. Domestic remittances became significant in improving rural household expenditure especially for the rural poor.

The impact of education and occupation choices on the falling urban-rural differences in wages was studied by Hnatkovska and Lahiri (2013) in the context of India between 1983 and 2010. Their findings suggest that almost 40 per cent of the wage convergence observed between urban and rural India was explained by converging individual characteristics such as education and occupation choices. Himaz and Aturupane (2011) applied a quantile regression to identify the importance of education on household welfare in Sri Lanka using five cross-section datasets between 1985 and 2006. Their paper found that people in higher quantiles who have greater consumption expenditure are more likely to have higher levels of education and better skills that complement education, thus enabling them to earn higher returns to education. The findings also indicated that residing in a rural area had a negative impact on the returns to education compared to residing in an urban area, especially at the top end of the welfare distribution. Finally, Sicular et al. (2007) examined the urban-rural gap in China and found that with better infrastructure and employment opportunities, people in rural areas can easily move to urban areas for employment. The paper noted that the exclusion of migrants and ignoring spatial price differentials across regions led to an over-estimation of the urban-rural gap.

3.2 Methods used in the literature

Measuring inequality is not straightforward. Many methods have been advanced to try to decompose inequality in order to better understand its causes. The Gini and Theil coefficients, for example, have been used often to decompose inequality into a within- and between-group component – how much of the inequality is due to inter-group effects and how much of it is due to intra-group effects. Although this helps understand the sources of inequality growth/decline, the between versus within decomposition does not identify the factors affecting the welfare or income distributions. Methods allowing the entire conditional distribution (rather than just the conditional variance) to be estimated were introduced (such as Machado and Mata, 2005). This technique creates a counterfactual distribution for one of the two groups

(rural, in this case) and compares it to the actual distributions in order to separate the urban-rural differences in welfare into two components – the first is the contribution of the differences in urban-rural household characteristics to the welfare gap (covariate effect; for example, the different education levels between the areas) and the second is how the differences in urban-rural returns to those characteristics contribute to the welfare gap (returns effect; for example, the returns to education). The decomposition gives a better understanding of how the contributions of characteristics and returns to characteristics have changed over time in affecting the urban-rural welfare gap. Nguyen et al. (2007) implemented this method in Vietnam in the period of rapid growth and rising inequality, noting that differences in covariates explain most of the urban-rural expenditure gap at lower quantiles; but for the rest of the expenditure distribution, the gap was primarily due to urban-rural differences in the returns to covariates. However, the drawback of this approach is that the decomposition is not detailed enough to compute the sub-components of the covariate effect; that is, to identify the specific sources that give rise to the differences in covariate distributions between the urban and rural areas.

Firpo et al. (2009) introduced a new technique which is an extension of the Machado-Mata (2005) decomposition identifying the detailed components of both, the returns and covariate effects. This is done through the estimation of (re-centred) influence function (RIF, hereafter) regressions. In this context, the RIF can be regarded as an unconditional quantile regression. Instead of using the traditional conditional quantile regression, this technique used the unconditional quantile regression. The (conditional) quantile regression technique was introduced by Koenker and Bassett (1978). This method estimates the effects of each explanatory variable on the entire distribution of expenditure. However, it is restrictive since a change in the distribution of covariates could change the interpretation of the estimated coefficients (Firpo et al., 2009). Fortin (2008) and Firpo et al. (2009) estimated the effect of union status on log wages of men in the United States and found large differences between the results using the conditional and unconditional quantile regressions. The estimates from the unconditional quantile regressions suggested that unionization progressively increases wages at the bottom end of the distribution, and reduces wages at the top end of the distribution which precisely explained the U-shaped changes observed in the actual wage data. The conditional regression results, in contrast, suggest that unionization

has a positive yet monotonically declining effect on wages without taking into account the observed changing pattern of the wage distribution. It is clear that the two different methods interpret results differently. Therefore, for the purpose of this study where it is important to understand how each household characteristic contributes to the welfare gap, an unconditional quantile regression is suitable.

In order to analyse the welfare gap between urban and rural areas, the conventional methodology proposed by Blinder (1973) and Oaxaca (1973) can be implemented. This standard decomposition stems from the notion that differences in expenditure between urban and rural households may arise from three possible sources – differences in endowments, differences in returns to endowments, and differences in unobservable characteristics. However, the Blinder-Oaxaca decomposition technique is carried out at the mean of the expenditure distribution. For the analysis of the urban-rural gap, it is important to examine the entire distribution. Firpo et al. (2009) apply a variant of the Blinder-Oaxaca decomposition to the estimates obtained from the unconditional quantile regression. This method was used by Thu Le and Booth (2014) in their study of the urban-rural welfare gap in Vietnam and will be used in this chapter.

This chapter contributes to the existing literature in two ways. Firstly, it explores the welfare gap between urban and rural households in Sri Lanka during a period of rapid growth. This has not been examined to date - Himaz and Aturupane (2011) observed the effect of education on household welfare between 1985 and 2006 in Sri Lanka, however urban and rural households were pooled together. This chapter extends the analysis to 2009/10 and isolates urban and rural households to identify welfare differences across the expenditure distribution and over time. Secondly from a methodological viewpoint, this chapter adds a new dimension with the use of an unconditional quantile regression as opposed to the conventional quantile regression technique. The use of a decomposition technique allows a further examination of the urban-rural gap to identify which factors were crucial in changing the welfare gap over time and across the expenditure distribution. It thus extends the existing literature with a detailed analysis of the contributing factors to urban-rural welfare gaps in Sri Lanka between 2002 and 2009/10. The next section will explain the data used in the chapter.

4 Data

4.1 Data and sample

The Household Income and Expenditure Surveys (HIES) of 2002 and 2009/10 are used in this study to cover the period of dramatic change in Sri Lanka. Data was collected in twelve consecutive monthly rounds in order to capture seasonal variations in income and consumption patterns. The 2002 survey was conducted from January 2002 through December 2002 and includes 16,920 households, and the 2009/10 survey was conducted from July 2009 through June 2010 and includes 17,182 households. For comparability, only provinces/districts included in both waves are included in the analysis – 17 out of the 25 districts in the country have been surveyed in both years. In the HIES carried out in 2002, the Northern and Eastern provinces were excluded because of the ongoing war in these areas. By 2009/10, data collection commenced in 5 additional districts in the aforementioned provinces. However, the districts were not surveyed for the entire 12 months - the Vavuniya district in the Northern province and the entire Eastern province were surveyed for 10 out of the 12 months whereas the Jaffna district in the Northern province was surveyed for 7 months. The remaining 3 districts in the country were left out of the 2009/10 survey due to ongoing resettlement activities.

The excluded districts include 2,776 households in the 2009/10 survey data, 40 per cent of whom are from urban areas. The descriptive statistics for the excluded districts in 2009/10 suggest that these districts have lower real expenditure per capita at every quantile, on average in comparison to the rest of the country – particularly at higher quantiles where large urban-rural differences in expenditure are seen). Households are smaller and these districts are predominantly rural (see Appendix 2). Descriptive statistics on these excluded districts obtained from the HIES 2009/10 are discussed in Appendix 1.

The data sets contain information on several aspects of living standards, including household and individual demographics, education, health, migration, employment, income, and expenditure on food and non-food. The HIE surveys record remittance

income transferred within the country, or received from abroad. The chapter uses adult non-student characteristics at the household level. This category is defined as people aged between 15 and 65 who are not in education at the time of the survey[33].

To compare the differences across urban and rural living standards, monthly household per capita expenditure was used – at 2002 prices. For the purpose of this study, the inclusion of private consumption (such as home-produced food) is an important indicator of economic welfare. Consumption data also reduces the issue of variability that income may have across the survey period (Deaton, 1997). Total household expenditure was calculated as the sum of expenditure on food and non-food items. Food expenditure includes the expenditure on purchased products as well as home-produced items. For home-produced items, the total quantity consumed was multiplied by the unit value if it were purchased in the market[34]. Non-food expenditure includes expenditure on housing[35], education, health, clothing, entertainment, communication, personal health care, household goods, transportation and vehicle maintenance.

Once we obtain total monthly expenditure (on food and non-food items) for a household, we then adjust for spatial variations in prices across the 17 districts. The price indices for both years (2002 and 2009/10) are presented in Appendix 2 (Table A2). According to the Department of Census and Statistics, Laspeyres price indices were calculated using unit prices of the typical consumer food basket for each district. Note that the price indices are observed at a given point in time at the district level, and are updated over the survey periods. By adjusting monthly expenditure to account for spatial price differences, we obtain expenditure values that are free from commodity price differences across districts. We are unable to disaggregate the indices further than district-level in order to observe price variations across urban and rural areas. However as shown in Table A2, the districts are either predominantly

[33] It is not conditioned on whether individuals are working or not since it is important to include the contributions of those involved in household work and other paid/unpaid family activities to consumption expenditure

[34] The HIES survey report explains that unit values were estimated according to the market value, including homegrown food or items received free of charge. This information is initially recorded in a separate form by the respondent (under the guidance of the enumerator during the visit to the household). It is later edited if necessary and recorded in the survey schedule by the enumerator.

[35] Rent values were imputed for those living in their own house

urban or rural – thus, the spatial price indices capture a part of the urban-rural price differences. Next, the monthly household expenditure is converted into real terms. This is useful for the comparison of expenditure over time. In order to do this, household expenditure is divided by the Consumer Price Index (base year is 2002). Thus, the expenditure values are at 2002 prices. Finally to obtain per-capita values, the real expenditure is divided by the number of members in the household. The main variable of the analysis is obtained - real per capita expenditure (RPCE). In this manner, we are able to account for spatial price variations across regions, which Sicular et al. (2007) noted to be important when estimating the urban-rural gap.

We identify a caveat in the analysis as the data does not include information on migration – we are unable to identify individuals who may have migrated from rural to urban areas due to improved infrastructure, for example. However, we attempt to deal with this issue by including controls for remittance flows received from abroad or within the country. Workers' remittances have become an increasingly attractive source of financing over the past three decades in Sri Lanka. In 2010, remittances accounted for 8.3 per cent of GDP (World Banka data), which is high compared to countries of broadly equal size and other Asian countries in 2005 (Lueth and Ruiz-Arranz, 2007). This finding was confirmed by Himaz and Aturupane (2011) – the paper stated that the remittance flows across the country and from foreign countries increased by nearly eight-fold in real terms over the 20 years that were analysed – being 3 per cent of consumption in 1985 and 13 per cent in 2006.

The next section explores descriptive statistics in order to understand some of the key movements across urban and rural sectors, over time, and across rich and poor households.

4.2 Descriptive statistics

Table 2 gives the within-quintile means for the main variables of interest by "real per capita expenditure" (RPCE) for urban and rural households[36].

From Table 2, the *real per capita expenditure* (RPCE) at any point in the distribution and over time is higher for urban households relative to rural households. However the gap between urban and rural expenditure per capita appears to be less pronounced at lower quintiles compared to higher quintiles in a given year, and in 2009/10 compared to 2002. Most of the explanatory variables focus on adult non-students in the household, defined as individuals between the ages of 15 and 65 years but are not in education.

Household demographic characteristics include the number of individuals living in the household (*household size*), the proportion of children below the age of 15 in the household (*proportion of children),* the average age of the adult non-students and the age squared divided by 100 (*average age, age squared*), proportion of women and Sinhalese[37] adult non-students (*proportion of women/Sinhalese*) in the household. *Ethnic* minorities are rarer in the rural sector in comparison to the urban sector. *Household size* has a negative relationship with per capita expenditure implying that larger households are poorer, on average. However, due to size economies this could be biased upwards. Poorer households spend a greater proportion of their income on rival goods such as food. Yet their consumption of certain items such as clothing, housing, water taps, etc. are shared among several members in the household – such bulk purchases suggest that the cost per person is lower (with a given standard of living) when more individuals live together, rather than separately. A common belief is that larger households tend to be poorer. However, a paper by Lanjouw and Ravallion (1995) discuss that this relationship between household size and

[36] Rural and estate households have been pooled together for the purpose of comparison, and since the estate households have similar behaviour/characteristics to rural households.
Quintiles are defined in two ways – in this section which looks at the mean values, the quintiles are defined as the bottom 20%, second 20%, middle 20%, fourth 20% and top 20% of the urban or rural sample based on the household RPCE. Moving on to the quantile regression, quintiles are defined as a point in the expenditure distribution, rather than the number of people falling into a certain category.
[37] Sinhalese form the biggest ethnic group; hence it will be taken as the default category.

poverty/expenditure may vanish at a certain point due to size economies in consumption. The paper employed a method whereby the total expenditure is divided by the household size raised to a power less than one (given a value of θ). This value is known as the size elasticity. As the value of θ decreases, consumption expenditure and household size become statistically independent. At values of θ larger than the threshold value, bigger households tend to have lower expenditure. For Sri Lanka, the value of θ at which point the household size and expenditure begin to have the negative relationship is at 0.3 for urban households and 0.6 for rural households. The low value of θ for urban households suggests that there is a negative relationship between household size and expenditure; whereas for rural households, the correlation between poverty and household size may vanish when there are size economies in consumption.

Human capital is measured by the average number of years of education acquired by the adult non-students (*education*). The average education level increases across quintiles as expected for urban and rural sectors, and at any given point, it is higher for urban households than for rural households. Dummy variables are used to identify whether or not the household received remittance income from within Sri Lanka (*local remit*) or from abroad (*foreign remit*) during the year. At a given point in time, a greater proportion of rural households receive local remittance income whereas a greater proportion of urban households receive foreign remittance income. Both, foreign and local remittance transfers have increased over time – specifically foreign remittance income received by urban households and local remittance income received by rural households.

The type of employment (*wage*, *self-employed or not employed*) is accounted for by the variables that measure the proportion of adult non-students in the household who are working in wage employment, self-employment or are not in the labour force[38]. We use the reported income by individuals to distinguish between the wage employees and self-employed. Individuals with more than one source of income were distinguished as being either wage/self-employed by looking at the source which

[38] This category is for the adult non-students who are unemployed, retired/disabled and stay-at-home parents during the sample period.

yielded the highest income. Looking at the descriptive statistics, there are a greater proportion of adults in some form of employment in the rural sector compared to the urban sector; this is true at any given point in the expenditure distribution and for both years. Both, being in wage and self-employment have a positive relationship with per capita expenditure across the distribution. Over time, the proportion of adults in urban areas working in self-employment has risen for the top end of the distribution and the proportion of adults in wage employment has fallen. In rural areas, the proportion of adults in wage employment has fallen over time whereas the proportion in self-employment has risen marginally.

Additional variables are used to account for the sector (*agriculture*, *manufacturing or services*) of the working adult non-students (that is, those in wage and self-employment) shown as proportions in the relevant form of employment (for example, the proportion of working adult non-students in the services industry). Per capita expenditure appears to have a negative relationship with agriculture employment and a positive relationship with service employment in both, urban and rural areas. Further, there is a larger proportion of agricultural workers and a smaller proportion of service sector workers in rural areas compared to urban areas. These findings are what would be expected in a developing country (Nguyen et al, 2007; Thu Le and Booth, 2014).

Having observed various changes in household characteristics, three interesting questions emerge. Firstly, what factors affect the urban and rural expenditure per capita, and how has it changed over time? Secondly, to what extent is household expenditure per capita determined by the observed productivity-related characteristics (as mentioned in this section) in urban and rural areas? Finally, how much of the urban-rural expenditure differential can be contributable to urban-rural differences in average characteristics, and how much of the expenditure differential can be contributable to the difference in returns to those characteristics and other factors which are not captured in the model? The chapter proceeds as follows. Section 5 will look at the determinants of urban and rural expenditure per capita across the expenditure distribution in both years; in section 6, the results obtained will be used in a decomposition that will enable isolation of the factors that give rise to the urban-rural gap in expenditure.

Table 2: Within-quintile means of variables by log RPCE for urban and rural households

2002	Urban (3,240 households)					Rural (13,680 households)				
Variable	**1st**	**2nd**	**3rd**	**4th**	**5th**	**1st**	**2nd**	**3rd**	**4th**	**5th**
RPCE (at 2002 prices)	1497	2420	3506	5315	11639	1052	1562	2096	2945	6648
Household size	5.67	5.09	4.54	4.17	3.62	5.09	4.59	4.22	3.92	3.67
Children (proportion)	0.25	0.20	0.17	0.16	0.12	0.26	0.21	0.19	0.17	0.15
Average age	30	33	35	37	41	31	33	34	36	38
Education	4.72	6.97	7.67	8.85	10.08	4.64	5.50	6.04	6.91	8.63
Women (prop.)	0.50	0.50	0.52	0.51	0.52	0.49	0.49	0.50	0.50	0.50
Sinhalese (prop.)	0.56	0.65	0.71	0.74	0.78	0.79	0.80	0.81	0.87	0.92
Local remit (0,1)	0.04	0.03	0.05	0.05	0.05	0.06	0.06	0.05	0.06	0.05
Foreign remit (0,1)	0.05	0.06	0.10	0.07	0.08	0.03	0.04	0.04	0.05	0.05
Wage (prop.)	0.24	0.26	0.27	0.31	0.37	0.26	0.28	0.29	0.27	0.33
Self-employed (prop.)	0.08	0.11	0.11	0.11	0.11	0.11	0.14	0.16	0.18	0.17
Not employed (prop.)	0.68	0.63	0.62	0.58	0.52	0.63	0.58	0.55	0.55	0.50
Of the employed:										
Agriculture (prop.)	0.09	0.06	0.06	0.05	0.02	0.49	0.47	0.41	0.32	0.20
Manufacture (prop.)	0.28	0.21	0.20	0.22	0.18	0.19	0.19	0.20	0.21	0.19
Services (prop.)	0.63	0.73	0.74	0.73	0.80	0.32	0.34	0.39	0.47	0.61

2009/10	Urban (4,192 households)					Rural (12,990 households)				
Variable	**1st**	**2nd**	**3rd**	**4th**	**5th**	**1st**	**2nd**	**3rd**	**4th**	**5th**
RPCE (at 2002 prices)	1788	2817	3884	5465	12576	1453	2171	2876	3931	8060
Household size	5.57	4.72	4.36	4.05	3.49	5.01	4.44	4.16	3.93	3.55
Children (proportion)	0.22	0.18	0.17	0.15	0.11	0.21	0.19	0.17	0.16	0.13
Average age	32	34	36	39	43	32	34	37	37	40
Education	5.78	6.72	7.50	8.41	9.72	4.72	5.63	6.36	7.21	8.76
Women (prop.)	0.53	0.53	0.54	0.53	0.53	0.50	0.51	0.52	0.52	0.52
Sinhalese (prop.)	0.58	0.64	0.69	0.72	0.76	0.66	0.72	0.78	0.84	0.89
Local remit (0,1)	0.05	0.04	0.05	0.05	0.02	0.11	0.09	0.08	0.08	0.07
Foreign remit (0,1)	0.08	0.08	0.09	0.10	0.12	0.06	0.04	0.06	0.05	0.08
Wage (prop.)	0.24	0.24	0.26	0.28	0.31	0.24	0.26	0.26	0.25	0.30
Self-employed (prop.)	0.07	0.11	0.12	0.13	0.14	0.12	0.15	0.16	0.18	0.19
Not employed (prop.)	0.69	0.65	0.62	0.59	0.55	0.64	0.59	0.58	0.57	0.51
Of the employed:										
Agriculture (prop.)	0.11	0.09	0.08	0.06	0.07	0.49	0.43	0.38	0.29	0.20
Manufacture (prop.)	0.29	0.26	0.25	0.23	0.16	0.22	0.23	0.24	0.22	0.19
Services (prop.)	0.60	0.65	0.67	0.71	0.77	0.29	0.34	0.38	0.49	0.61

The key variable of interest is "real per capita expenditure", calculated in Sri Lankan rupees at 2002 prices (1 USD ≈ 146 LKR). The explanatory variables are computed using adult non-student characteristics. Education is measured as the average years of education obtained by adult non-students. The employment variables give the proportion of adult non-students in wage or self-employment, or not a part of the work force (not employed). The category of not being in employment is the adult non-students who are unemployed, retired/disabled and stay-at-home parents during the sample period. For the wage and self-employed adults, the sectoral variables capture the proportion of wage/self-employed adults in agriculture, manufacturing or the services sector

5 Determinants of urban and rural per capita expenditure

5.1 Method

The chapter focuses on the link between household characteristics and real per capita expenditure (RPCE) across urban and rural areas. As per capita expenditure varies across the distribution in both, urban and rural areas, there is a need to examine the entire distribution of expenditure rather than simply focussing on the mean (as an Ordinary Least Squares estimation would do). For this purpose, a quantile regression is more suitable.

The unconditional quantile regression technique by Firpo et al. (2009)[39] estimates the impact of changes in the distribution of covariates (the explanatory variables) on the unconditional marginal distribution of log RPCE (the outcome variable). According to this method, a regression of the estimated re-centred influence function (RIF) on a set of explanatory variables is run. This approach of RIF uses the concept of the influence function. The influence function of a distributional statistic $v(y)$ represents the influence an observation has on the distributional statistic. Adding the distributional statistic $v(y)$ back to the influence function gives the RIF:

(1) $$RIF(y; v) = v + IF(y; v)$$

The usual outcome variable Y (representing the natural logarithm of RPCE in this chapter) in the regressions will be replaced by the re-centred influence function $RIF(y; v)$ of the statistic *v*. In the case of the mean, the influence function is: $IF(y; \mu) = y - \mu$

[39] Firpo et al. (2009) explain the method in greater detail. For the purpose of this chapter, a brief summary of the method is presented here. The regressions were run using the user-written STATA command *rifreg* by Firpo et al. (2009)

The RIF becomes:

$$RIF(y;\mu) = \mu + (y - \mu) = y \quad (2)$$

Since the $RIF(y;\mu)$ simply reduces down to the outcome variable y, the RIF regression for the mean corresponds to the standard Ordinary Least Squares (hereafter, OLS) wage regression; that is, if the mean is the statistic of interest, the estimation of RIF-OLS for the mean will become exactly the OLS estimation.

Similarly, the influence function can be computed for various quantiles of the expenditure distribution. At the quantile θ, the RIF will be:

$$RIF(y;q_\theta) = q_\theta + IF(Y;q_\theta) = q_\theta + \frac{\theta - 1\{y \le q_\theta\}}{f_y(q_\theta)} \quad (3)$$

This new method used by Firpo et al. (2009) that estimates the unconditional quantile regression can be done through one of the three techniques: Ordinary Least Squares (namely, RIF-OLS), logistic (namely, RIF-logit) or non-parametric (namely, RIF-nonparametric). This chapter uses RIF-OLS for simplicity[40]. The RIF estimations are then included in the regressions instead of log RPCE which will be explored in the next section.

5.2 Model specifications

To understand the relationship between the natural log of real per capita expenditure (RPCE) and various household characteristics, especially how they differ across urban and rural areas over the entire distribution of log RPCE, quantile regressions of the following form will be estimated:

$$Y_i = \alpha + \beta X_i + \gamma U_i + \delta U_i . X_i + \varepsilon_i \quad (4)$$

[40] Firpo et al. (2009) obtained similar results using all three estimation techniques

where "Y" is the dependent variable – log RPCE of household i, U_i is the urban dummy, X_i is the vector of explanatory variables for the household (excluding "u"), and $U_i.X_i$ is the interaction between the urban dummy and explanatory variables. The vector of coefficients β represents the returns to characteristics; γ and δ coefficients are the respective intercept and slope differential for the urban dummy variable. The explanatory variables (X_i) include education, demographic, employment and geographical characteristics of the household of adult non-students. To control for regional differences, dummies for regions will be included. The way each characteristic is captured was explained in Section 4.2.

This chapter analyses the urban-rural gap in welfare in two stages. Firstly, the urban and rural households will be analysed separately to observe the factors affecting real per capita expenditure for both these sectors. The second part will be a decomposition of the urban-rural gap in the expenditure distribution. The total urban-rural gap can be disaggregated into two components. The first is the contribution of urban-rural differences in the distributions of covariates such as education to the urban-rural gap (covariate effect). The second component arises from the differences in the distributions of returns to these covariates (returns effect). To decompose the urban-rural gap across the log RPCE distribution and isolate the two effects, the results obtained from the quantile regression are applied to a Blinder-Oaxaca decomposition. This technique will be discussed in greater detail in **Section 6** after estimating the quantile regressions to identify the factors affecting urban and rural real per capita expenditure across the distribution.

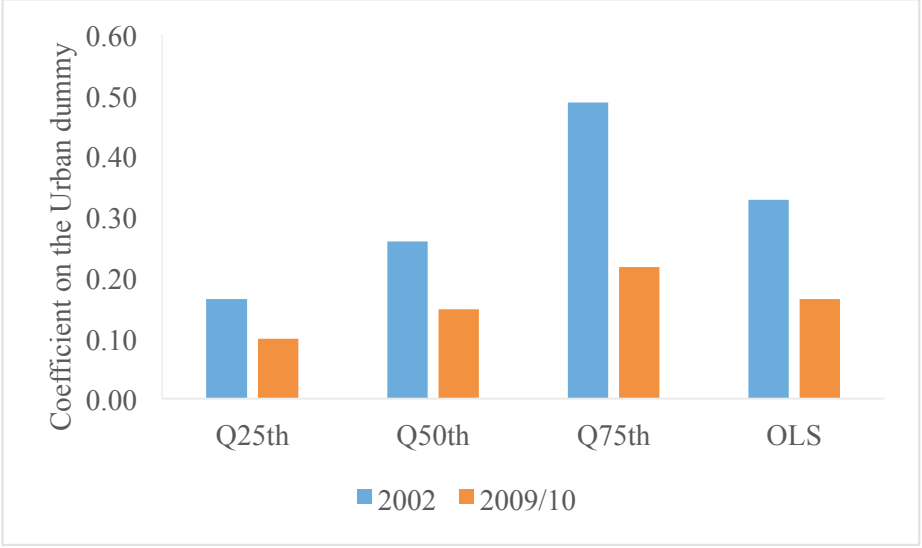
5.3 Results

In this section, the household level determinants of real per capita expenditure are analysed in urban and rural areas for both sample periods – 2002 and 2009/10. We start by testing whether the urban and rural samples need to be analysed in isolation. In order to do this, quantile regressions were estimated for the pooled sample (urban and rural) to test and confirm that there are significant differences in per capita expenditure across these two areas which are worth analysing further. The two tests that were conducted will be explained in detail below.

As noted from the descriptive statistics in Section 4.2, a "raw" gap in per capita expenditure was identified between urban and rural households. A regression allows us to control for several household factors and identify the "pure" gap which may or may not exist after controlling for other factors. The first test employs a model of Equation 4 that includes the intercept, urban dummy and the explanatory variables run at the mean using OLS and at selected quantiles using the unconditional quantile regression for the entire sample. The quantile regression allows for returns to vary with the households' positions in the distribution that are not accounted for by a mean regression. The inclusion of an urban dummy in a regression identifies the urban-rural gap in per capita expenditure as the "pure gap" after controlling for other factors.

Figure 2 plots the coefficients on the urban dummy variables in a bar chart[41]. A detailed table of estimation results including household characteristics is given in the Appendix (Table A3). The coefficients are positive and significant suggesting that, other things being equal, an urban household has higher per capita expenditure than a comparable rural household. Between 2002 and 2009/10, the urban-rural gap has fallen; the coefficients on the urban dummies are lower in 2009/10 compared to the coefficients in 2002. In 2002, urban households spent 23 per cent more than their rural counterparts at the median. By 2009/10 this has dropped; urban households spent 14 per cent more than their rural counterparts (at the median).

Figure 2: Quantile regression for full sample (coefficients on urban dummies)



[41] Here we interpret the coefficient as (exp(β)-1) which is the exponential of the coefficient (β).

Looking at the entire distribution at a given point in time, the coefficients on the urban dummies are increasing monotonically up the expenditure distribution which suggests that the urban-rural differences in per capita expenditure are higher for households at the top end of the distribution. This is true for 2002 as well as 2009/10. In 2002 for the 25th quantile, households in urban areas spent 16 per cent more than their rural counterparts. At the other end of the distribution at the 75th quantile, households in urban areas spent almost 50 per cent more than their rural counterparts. By 2009/10, the coefficients on the urban dummy variables are much lower at any given quantile, but are rising (from a 10 per cent expenditure differential at the bottom end to a 22 per cent differential at the top end of the distribution). This suggests that the per capita expenditure differences between urban and rural areas are larger for richer households.

Having analysed the movements of the urban dummy across the distribution and over time, the above test confirms that there are indeed significant urban-rural differences in per capita expenditure at the 1 per cent significance level. In the next test we estimate Equation 4, including all variables being interacted with the urban dummy variable. The estimation results from this specification are given in the Appendix (Table A4). An F-test was carried out to test the hypothesis that all the coefficients of the interaction terms (between the urban dummy and the observables) are equal to zero. For both survey years, the test rejects the null hypothesis[42] – therefore, there are significant differences in the returns to household characteristics between urban and rural areas. The specification includes the intercept, urban dummy, the explanatory variables and the interaction terms of the urban dummy with the set of explanatory variables at the mean using OLS and at selected quantiles using the unconditional quantile regression framework. The interaction terms where the urban dummy is interacted with each explanatory variable identify urban-rural differences in the coefficients. Therefore, we can directly test whether the *returns* to household characteristics (given by the β terms in equation 4) differ between the two sectors rather than solely focussing on the household characteristics (given by the X terms in equation 4). This has been used by Nguyen et al. (2007) and Thu Le and Booth (2014)

[42] The null hypothesis can be rejected at the 1 per cent significance level at the mean and the 25th, 50th and 75th quantiles. The P-value is 0.00 in all cases

in the case of Vietnam and is a useful way of identifying the most important factors affecting the urban-rural gap[43]. The coefficients on the urban dummy measure the urban-rural gap that is not explained by the covariates in the regressions. We note the insignificance of the urban dummy in most of the quantiles for both years (with the exception of the means in both years where the urban coefficient is significant at the 1 per cent level).

Having confirmed that urban-rural differences do exist even after controlling for household characteristics and that there are significant differences in returns to household characteristics between urban and rural areas, we move to the main part of the analysis. This involves examining the determinants of per capita expenditure at selected quantiles using the unconditional quantile regression for urban and rural sectors separately. The estimation results are reported in Tables 3 and 4.

[43] The assumption made about the distribution of the error terms is different in comparison to using separate urban and rural samples – the use of interaction terms in a pooled sample assumes that the error terms are drawn out from the same error distribution. However, if the sample is split between urban and rural households, the error terms are different since they are drawn out from two separate samples.

Table 3: Determinants of urban household expenditure per capita at the mean and selected quantiles in 2002 and 2009/10

Variables	Q25th		Q50th		Q75th		OLS	
	2002	**2009/10**	**2002**	**2009/10**	**2002**	**2009/10**	**2002**	**2009/10**
Personal characteristics (average of non-student adults):								
Female	-0.017	-0.001	0.040	-0.087	-0.150*	-0.055	-0.068	-0.064
	(-0.27)	(-0.22)	(0.55)	(-1.63)	(-1.84)	(-0.82)	(-1.39)	(-1.57)
Sinhalese	0.138***	0.058**	0.065*	0.057**	-0.019	0.008	0.067***	0.023
	(4.22)	(2.06)	(1.85)	(2.19)	(-0.52)	(0.27)	(2.77)	(1.16)
Average age	-0.038***	-0.023***	-0.059***	-0.028***	-0.047***	-0.039***	-0.048***	0.030***
	(-7.32)	(-5.97)	(-10.25)	(-6.87)	(-7.34)	(-7.94)	(-12.43)	(-10.01)
Average age^2	0.045***	0.023***	0.077***	0.029***	0.067***	0.042***	0.064***	0.031***
	(7.69)	(5.18)	(11.51)	(6.44)	(8.88)	(7.23)	(14.02)	(9.14)
Education	0.116***	0.090***	0.120***	0.109***	0.137***	0.125***	0.113***	0.107***
	(19.97)	(19.77)	(24.98)	(25.93)	(21.48)	(23.99)	(31.86)	(31.06)
Job type (average of non-student adults; default – unemployed/not employed):								
Wage	0.029	0.052	0.371***	0.132**	0.465***	0.228***	0.389***	0.121***
	(0.44)	(0.92)	(4.72)	(2.25)	(5.15)	(3.05)	(6.84)	(2.68)
Self-employed	0.383***	0.370***	0.452***	0.378***	0.544***	0.397***	0.516***	0.353***
	(4.63)	(5.98)	(4.32)	(5.17)	(4.72)	(4.18)	(7.50)	(6.45)
Job sector (average of non-student adults; default - agriculture):								
Manufacturing	-0.076	-0.013	0.030	-0.052	0.057	-0.069*	-0.017	-0.049*
	(-1.32)	(-0.35)	(0.59)	(-1.46)	(1.11)	(-1.74)	(-0.48)	(-1.87)
Services	0.140***	0.152***	0.171***	0.169***	0.180***	0.181***	0.161***	0.166***
	(2.89)	(3.91)	(3.31)	(4.22)	(3.50)	(5.77)	(4.53)	(5.03)
Other demographic characteristics:								
Children (proportion)	0.164*	0.229***	0.592***	0.249***	0.632***	0.279***	0.424***	0.234***
	(1.68)	(2.93)	(5.76)	(3.26)	(5.94)	(3.23)	(6.12)	(4.13)

Household size	-0.381***	-0.462***	-0.497***	-0.478***	-0.468***	-0.498***	-0.457***	-0.487***
(log)	(-12.09)	(-16.97)	(-14.61)	(-18.15)	(-12.67)	(-14.74)	(-18.89)	(-22.99)
Foreign remit	0.308***	0.240***	0.264***	0.240***	0.209***	0.223***	0.237***	0.241***
(0,1)	(6.52)	(6.60)	(4.46)	(6.39)	(3.35)	(4.71)	(6.03)	(8.05)
Local remit	0.215***	0.056	0.231***	0.033	0.252***	-0.085	0.222***	-0.028
(0,1)	(3.47)	(0.99)	(3.25)	(0.56)	(3.17)	(-1.36)	(4.63)	(-0.68)
Regions (default – Sabaragamuwa):								
Western	0.301***	0.325***	0.305***	0.291***	0.410***	0.339***	0.325***	0.307***
	(4.54)	(5.24)	(4.64)	(5.47)	(6.91)	(6.02)	(7.17)	(7.31)
Central	0.174***	0.234***	0.208***	0.154**	0.356***	0.162**	0.241***	0.157***
	(2.38)	(3.28)	(2.83)	(2.43)	(5.14)	(2.45)	(4.75)	(3.22)
Southern	0.027	0.323***	0.102	0.232***	0.216***	0.286***	0.104**	0.258***
	(0.40)	(4.88)	(1.38)	(4.17)	(3.20)	(4.81)	(2.03)	(5.92)
North West	0.139	0.240***	0.174*	0.220***	0.273***	0.185**	0.204***	0.193***
	(1.58)	(3.21)	(1.92)	(3.39)	(3.31)	(2.57)	(3.47)	(3.69)
North Central	0.201**	0.229***	0.084	0.235***	0.052	0.268***	0.103	0.249***
	(2.13)	(2.68)	(0.84)	(3.00)	(0.53)	(2.97)	(1.49)	(4.02)
Uva	0.045	0.234**	0.218**	0.170**	0.414***	0.102	0.141**	0.195***
	(0.46)	(2.55)	(2.18)	(2.08)	(3.78)	(1.21)	(2.05)	(3.02)
Constant	7.433***	7.815***	7.897***	8.207***	8.194***	8.731***	7.920***	8.326***
	(58.63)	(74.38)	(57.80)	(82.06)	(56.56)	(73.95)	(84.97)	(109.72)
Number of observations	3,240	4,192	3,240	4,192	3,240	4,192	3,240	4,192
R-squared	0.24	0.19	0.29	0.25	0.25	0.23	0.43	0.36

Refer to Table A8 for a detailed explanation of how the variables were constructed; *t* statistics in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; the standard errors are bootstrapped with 500 replications

Table 4: Determinants of rural household expenditure per capita at the mean and selected quantiles in 2002 and 2009/10

Variables	Q25th		Q50th		Q75th		OLS	
	2002	**2009/10**	**2002**	**2009/10**	**2002**	**2009/10**	**2002**	**2009/10**
Personal characteristics (average of non-student adults):								
Female	0.011	-0.003	-0.012	-0.027	-0.085**	-0.082**	-0.016	-0.041*
	(1.44)	(-0.12)	(-0.42)	(-0.99)	(-2.22)	(-2.31)	(-0.75)	(-1.90)
Sinhalese	-0.100***	0.003	-0.069***	0.003	-0.041**	-0.030	-0.077***	-0.012
	(-5.22)	(0.14)	(-3.71)	(0.19)	(-2.03)	(-1.58)	(-5.33)	(-0.92)
Average age	-0.004*	-0.006***	-0.016***	-0.015***	-0.022***	-0.024***	-0.013***	-0.015***
	(-1.76)	(-2.93)	(-7.23)	(-7.58)	(-7.56)	(-9.34)	(-7.76)	(-9.54)
Average age^2	0.004*	0.005**	0.019***	0.015***	0.027***	0.027***	0.016***	0.015***
	(1.73)	(2.38)	(7.40)	(6.47)	(7.96)	(8.68)	(8.16)	(8.39)
Education	0.064***	0.064***	0.090***	0.083***	0.124***	0.107***	0.097***	0.089***
	(29.31)	(27.35)	(40.78)	(37.56)	(41.60)	(36.97)	(53.01)	(48.04)
Job type (average of non-student adults; default – unemployed/not employed):								
Wage	0.008	0.027	0.021	0.055	0.087**	0.179***	0.087***	0.109***
	(0.33)	(0.99)	(0.76)	(0.80)	(2.40)	(5.01)	(4.03)	(5.01)
Self-employed	0.251***	0.200***	0.291***	0.191***	0.278***	0.247***	0.294***	0.223***
	(8.56)	(6.71)	(8.66)	(6.11)	(6.06)	(5.73)	(11.58)	(9.04)
Job sector (average of non-student adults; default - agriculture):								
Manufacturing	0.049**	0.018	0.085***	$-3.82\text{x}10^{-4}$	$3.49\text{x}10^{-5}$	-0.059***	0.033**	-0.029**
	(2.61)	(0.97)	(4.48)	(-0.02)	(0.00)	(-2.93)	(2.23)	(-2.15)
Services	0.119***	0.112***	0.194***	0.166***	0.233***	0.191***	0.162***	0.131***
	(8.17)	(7.29)	(12.94)	(11.38)	(12.47)	(10.41)	(14.01)	(11.45)
Other demographic characteristics:								
Children	0.063*	0.186***	0.119***	0.198***	0.313***	0.302***	0.221***	0.237***
(proportion)	(1.68)	(4.72)	(3.13)	(5.46)	(6.60)	(6.80)	(7.44)	(8.22)

Household size (log)	-0.361*** (-27.83)	-0.356*** (-24.81)	-0.467*** (-34.24)	-0.426*** (-30.98)	-0.489*** (-25.65)	-0.474*** (-25.70)	-0.443*** (-39.84)	-0.422*** (-37.25)
Foreign remit (0,1)	0.140*** (5.51)	0.136*** (5.66)	0.186*** (6.31)	0.199*** (8.36)	0.214*** (5.21)	0.300*** (9.38)	0.218*** (9.80)	0.236*** (12.56)
Local remit (0,1)	0.056** (2.25)	0.020 (1.31)	0.086*** (3.35)	0.053** (2.60)	0.131*** (3.78)	0.079*** (3.21)	0.103*** (5.21)	0.067*** (4.27)
Regions (default – Sabaragamuwa):								
Western	0.200*** (10.36)	0.202*** (9.70)	0.286*** (14.60)	0.230*** (11.61)	0.361*** (14.13)	0.285*** (11.97)	0.268*** (17.41)	0.221*** (14.15)
Central	0.109*** (5.31)	0.012 (0.49)	0.104*** (5.25)	0.026 (1.23)	0.071*** (3.06)	0.101*** (4.29)	0.089*** (5.67)	0.039** (2.30)
Southern	0.110*** (5.00)	0.123*** (5.46)	0.158*** (7.33)	0.152*** (7.36)	0.103*** (3.99)	0.163*** (6.93)	0.109*** (6.50)	0.128*** (7.99)
North West	0.071*** (3.27)	0.047* (1.93)	0.080*** (3.74)	0.044* (1.93)	0.044* (1.70)	0.039 (1.52)	0.056*** (3.33)	0.025 (1.44)
North Central	0.189*** (7.80)	0.129*** (4.96)	0.213*** (8.25)	0.155*** (6.24)	0.167*** (5.13)	0.181*** (5.98)	0.165*** (8.44)	0.135*** (7.01)
Uva	-0.019*** (-077)	-0.063** (-2.20)	0.026 (1.14)	-0.022 (-0.90)	0.035 (1.30)	0.031 (1.20)	0.023 (1.24)	-0.020 (-1.04)
Constant	7.283*** (157.16)	7.564*** (161.05)	7.702*** (161.30)	8.008*** (181.95)	7.974*** (129.69)	8.373*** (148.68)	7.655*** (208.19)	8.015*** (227.33)
Number of observations	13,680	12,990	13,680	12,990	13,680	12,990	13,680	12,990
R-squared	0.16	0.17	0.24	0.24	0.24	0.22	0.35	0.34

Refer to Table A8 for a detailed explanation of how the variables were constructed; *t* statistics in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; the standard errors are bootstrapped with 500 replications

Education of adults is positively related to household per capita expenditure in urban and rural areas, across the expenditure distribution, and in both survey periods. In 2002 at the median, an additional year of education increased household per capita expenditure by 9 and 12 per cent in rural and urban areas respectively. By 2010 at the median, the rural returns to education were 8.3 per cent and urban returns were 10.9 per cent. Further, the returns to education vary across the expenditure distribution; returns to education of the urban sector remain higher than returns to education of the rural sector. However, the returns between the two sectors became smaller across the distribution – in the 25th quantile in 2002, the urban returns to education were 11.6 per cent whereas rural returns to education were 6.6 per cent; by the 75th quantile, the urban returns to education were 13.7 per cent and rural returns were 12.4 per cent.

Next, consider the *employment type.* Households with adults working in self-employment consistently have higher per capita expenditure than comparable households with adults working in wage employment or not employed/unemployed across the expenditure distribution. Not being in the labour force has the greatest negative association with per capita expenditure. This is true for both, urban and rural areas. At any given point in the distribution, the returns to self-employment are higher for urban households compared to rural households. In 2002, the returns to self-employment in comparison to the returns to wage employment at the median were 45.2 per cent for urban households and 29.1 per cent for rural households. In 2010 at the median, the urban returns were 37.8 per cent and rural returns were 19.1 per cent. At a given point in time, the returns to self-employment rose across the expenditure distribution – with the exceptions of urban households in 2009/10 and rural households in 2002 when the returns remained fairly flat across the distribution. The fact that self-employment pays higher returns than wage employment could come across as an unusual finding. It can be explained in the following way. Parker (2009) stressed that wage returns differ from returns to self-employment for several reasons – firstly, it is difficult to interpret the salary of a self-employed individual as he/she chooses to pay this for him/herself; secondly, the returns to self-employment are not purely the return to labour, but also include the return to capital. Hence this finding must be treated with caution.

Job sectors must be considered to further examine the changing returns to employment. Households with adults working in agriculture have lower per capita expenditure relative to households working in services or manufacturing industries. This is in line with the finding observed by Thu Le and Booth (2014) in Vietnam. Rural and urban households where adults are working in services had the highest per capita expenditure – at the median, the returns to working in services was approximately 17 per cent higher compared to working in agriculture in both years for rural and urban households. For rural households, the returns from working in the service industry increased across quantiles – from 12 per cent at the 25th quantile in 2002 to 23 per cent at the 75th quantile - suggesting that richer households got higher returns from employment in this industry. A similar pattern was observed in 2009/10. In urban areas, the coefficient on services does not vary across the distribution of expenditure – this is true for both years.

The proportion of children and the natural log of household size measured the effects of household *demographics*. The negative coefficient on household size suggests that bigger households have lower per capita expenditure. Thu Le and Booth (2014) found the same finding for Vietnam; the study also found that the proportion of children inversely linked to per capita expenditure – however, the opposite finding was observed for Sri Lanka where the presence of children in the household is linked with higher per capita expenditure. The coefficients on these two demographic variables become larger across the expenditure distribution suggesting that the negative link between house size and per capita expenditure, and the positive link between the presence of children and per capita expenditure are greater for richer households. As observed in Section 4.2, the negative relationship between household size and expenditure may possibly vanish at a size elasticity of 0.6 for rural households due to size economies of consumption. However for urban households, we observed a low value of the size elasticity (0.3) suggesting that larger households are likely to have less expenditure.

The receipt of foreign and local *remittances* is positively associated with urban and rural per capita expenditure in both years. Remittance flows were measured by a dummy variable which had a value of 1 for households that received foreign/local remittance income or a value of zero otherwise. Firstly, the impact of foreign

remittances on per capita expenditure is discussed. An urban (rural) household obtaining foreign remittance had approximately 25 (19) per cent higher per capita expenditure than their counterparts at the median of the expenditure distribution – the coefficients are similar in both years. For rural households, the positive relationship between foreign remittance and per capita expenditure rose greatly across the expenditure distribution especially in 2010 – this implies that receiving foreign remittance had a larger positive link with per capita expenditure for richer rural households in comparison to poorer households. In 2009/10 at the 25th quantile, rural households receiving foreign remittances had 14 per cent more per capita expenditure, and at the 75th quantile, they were 30 per cent better off compared to those who did not receive foreign remittances. Thu Le and Booth (2014) explained this in the following way – in order for an individual in the family to migrate to a foreign country for the purpose of working, the family will incur an initial cost (this cost could be for travel, setting up, acquiring skills, learning the language). For poorer households, this cost is likely to be covered by borrowing. Once the poor families start receiving foreign remittances, they will first have to pay the borrowed money back before they use the remittance receipt for their expenditure. Hence, a rich household could have a larger increase in their expenditure once they have received foreign remittances compared to poor households since they do not have to re-pay borrowed money. For urban households receiving foreign remittances, in 2002 the positive effect decreases across the expenditure distribution but in 2009/10 the effect is fairly stagnant across the distribution.

Domestic remittances also had a positive association with urban and rural per capita expenditure; however an exception was 2009/10 for urban households where local remittances shared no significant association (at the 10 per cent level) with per capita expenditure at any point in the distribution. In 2002 however, urban households receiving domestic remittances had 22 per cent more per capita expenditure compared to those who do not receive domestic remittances at the 25th quantile with the effect becoming slightly larger for richer households (25 per cent). For rural households in 2002, receiving domestic remittances had an increasingly positive association with per capita expenditure across the distribution suggesting that richer households further benefitted from receiving this form of income. However in 2009/10, local remittances had a significant and positive relationship with rural per capita expenditure at the

median and 75th quantile (at the 1 per cent significance level) but not at the 25th quantile. The coefficient on the domestic remittance variable is however always smaller than the coefficient of the foreign remittance variable; at the median in 2009/10, rural households receiving foreign remittances had 20 per cent higher per capita expenditure relative to those who did not, whereas rural households receiving local remittances saw a 5 per cent rise in per capita expenditure compared to those who did not receive this income. This infers that households receiving income from abroad are likely to have higher per capita expenditure compared to households receiving domestic transfers.

Finally, the *region* dummy variables give some noteworthy findings. The Western province has the highest living standards; this is true across urban and rural areas. This is consistent with the fact that the capital city, Colombo, is a part of the Western province. Colombo is the major economic centre in Sri Lanka and has the lowest incidence of poverty. Although the Western province is primarily urban, almost 30 per cent of population in this province is from the rural sector according the Census of Population and Housing (2011). Additionally, the findings show that the rural poor of the Uva province have the lowest standard of living with the exception of the 75th quantile.

The results obtained so far suggest that the return to working in certain types of employment, education, and other household demographics are significant indicators of urban and rural per capita expenditure. Further, regression estimations from the pooled sample suggested that the urban dummy is highly significant – therefore, there are urban-rural differences in per capita expenditure that can be explained by differences between the two areas due to household endowments and the differences in returns to endowments. Given the recent changes in the Sri Lankan economy during the period of analysis, including the end of a civil war, growth of industry and a reduction in poverty, our results are in line with these changes. Further, the importance of remittances, services sector and education are in line with findings in literature on other developing countries. The next section of the chapter will identify the contribution of each covariate used in the former regressions to the urban-rural expenditure gap, and how the composition of the gap has changed over time.

6 Factors contributing to the urban-rural gap in expenditure

6.1 Method

The urban-rural differences in the distribution of log RPCE can be due to two reasons – (1) differences in household characteristics between urban and rural areas, or (2) differences in the returns to those characteristics between urban and rural areas. As observed from the previous section, differences in returns do explain the differences in per capita expenditure between households in urban and rural areas – this section will look at the magnitude of the contribution and how it has changed between 2002 and 2009/10 which will help explain the changing pattern of the urban-rural welfare gap. More importantly, the impact of each observable characteristic (for example, education) or the returns to those characteristics (returns to education) on the urban-rural gap can be estimated. This is done by applying a decomposition technique to the estimation results obtained in the previous section for urban and rural households.

In order to carry out this decomposition, the Blinder-Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973) is used[44]. The form of the decomposition is as follows:

$$\hat{Y}_u - \hat{Y}_r = (\hat{X}_u - \hat{X}_r)\hat{\beta}_r + (\hat{\beta}_u - \hat{\beta}_r)\hat{X}_r + (\hat{X}_u - \hat{X}_r)(\hat{\beta}_u - \hat{\beta}_r) \quad (5)$$

where Y is the natural logarithm of real per capita expenditure (RPCE) for the household, hence $\hat{Y}_u - \hat{Y}_r$ is the difference in log RPCE between urban and rural areas. $\hat{X}$ is a vector of observed characteristics, $\hat{\beta}$ is a vector of estimated coefficients obtained from running the regression model of log RPCE on a set of explanatory variables and a constant. This is a three-fold decomposition, where the differences in log RPCE can be explained in three parts.

[44] The decomposition was run using the STATA command *oaxaca8* by Firpo et al. (2009) which they employed in their paper

The first part is given by:

$$(\hat{X}_u - \hat{X}_r)\hat{\beta}_r$$

which is the expenditure differential that is due to differences in household characteristics between urban and rural areas (the *endowment* effect).

The second part is given by:

$$(\hat{\beta}_u - \hat{\beta}_r)\hat{X}_r$$

which is the expenditure differential that is due to differences in returns to household characteristics between urban and rural areas (the *coefficient* effect). It looks at the contribution of differences in coefficients of the urban and rural household regressions (including differences in the intercepts) to the overall expenditure gap.

The third part is given by:

$$(\hat{X}_u - \hat{X}_r)(\hat{\beta}_u - \hat{\beta}_r)$$

which is an interaction term accounting for the fact that there could be differences in endowments and coefficients existing simultaneously between the two groups.

The decomposition given by Equation 5 is estimated from the viewpoint of rural households; that is, the endowment effect measures the expected change for the group of rural households' mean expenditure per capita (or at a given quantile which will be explained further below) if rural households had urban characteristics. Similarly, the coefficient effect measures the expected change for the group of rural households' mean expenditure per capita (or at a given quantile) if rural households received returns of urban households[45].

[45] Alternatively, we can carry out a reverse three-fold decomposition to estimate the three effects from the viewpoint of urban households. However, the command "oaxaca8" is used for decomposition at different quantiles, rather than the "oaxaca" command for the decomposition at the mean – the former does not allow estimation of the reverse decomposition.

A limitation of the Blinder-Oaxaca decomposition for the purpose of analysing urban/rural expenditure differences across the entire distribution of expenditure is that it could only be applied to the mean. The unconditional quantile regression, however, as introduced by Firpo et al. (2009) obtains an estimate of the marginal impact of a unit of change in any explanatory variable on the unconditional quantiles of log RPCE (discussed in Section 5.1). This enables an application of the Blinder-Oaxaca decomposition directly to the estimation results obtained from the unconditional quantile regression.

There are two advantages of using this over other methods. Firstly, there is no need to run several simulations – as is the case with the conditional quantile regression decomposition put forward by Machado and Mata (2005). The technique by Machado and Mata (2005) estimates the entire conditional distribution F_x over a new distribution G_x in order to obtain a counterfactual unconditional distribution of Y for rural households (where Y is the main variable of interest, log real per capita expenditure). The counterfactual distribution is then compared with the empirical distributions for urban and rural households to identify the differences in characteristics and returns to characteristics that exist between the two groups. By contrast, RIF estimates the conditional distribution F_x only at one specific point in the distribution. Secondly, the use of the Blinder-Oaxaca decomposition allows us to isolate the contributions made by the differences in urban-rural returns and characteristics from *each* explanatory variable to the urban-rural expenditure gap at any given quantile along the distribution.

6.2 Results

Table 5 reports the urban-rural expenditure gap in 2002 and 2009/10, along with the contributing factors at the mean and selected quantiles. The total predicted gap at a particular point in time rose across the expenditure distribution suggesting that differences in per capita expenditure between urban and rural households are higher amongst richer households – in 2002 (2009/10), the overall expenditure gap between urban and rural households was 0.42 (0.25) log points at the 25^{th} quantile and rose to 0.60 (0.33) log points at the 75^{th} quantile. Over time, the expenditure gap has fallen

sharply – at the median in 2002, the predicted gap was 0.51 log points and it fell to 0.30 log points by 2009/10. This fall in the urban-rural welfare differences over time is in line with what we have observed in the Sri Lankan economy, including the fall in poverty (particularly rural poverty), high growth and the end of the war.

In both years, the urban-rural expenditure gap arose from two factors: the urban-rural differences in endowments (differences in X terms of the regression), and the differences in returns they get from those endowments[46] (coefficients; that is, differences in β terms). The welfare gap is reduced by the differences in endowments interacted with the differences in returns (interaction terms; that is, differences in $X*\beta$) – with one exception in 2002 (at the 25th quantile) where the interaction effect led to an increase in the welfare gap. In 2002, the expenditure gap can be explained more by the differences in returns rather than the differences in endowments at any point on the distribution. In 2010, this is true at the 25th and 50th quantiles; however at the 75th quantile, the differences in endowments and the differences in returns contribute equally to the overall expenditure gap. The differences in characteristics interacted with the differences in the returns to characteristics simultaneously (as given by the interaction term in the decomposition) reduced the welfare gap, especially for the richest households – an exception was noted in 2002 at the 25th quantile where the interaction effect *increased* the welfare gap by 12 per cent.

In order to identify the underlying factors that contributed towards the differences in endowments, returns and the simultaneous differences between the two former effects, some key household characteristics will be examined and discussed.

Considerable differences in *endowments* between urban and rural households that contribute towards the urban-rural expenditure gap will be discussed, namely education, regional differences, sector of employment and employment activity. Education was the largest contributing variable in both years – higher education levels by urban households compared to rural households explained over 60 per cent of the "endowment effect" at any point across the expenditure distributions in 2002; by

[46] The rural returns are matched with urban characteristics to observe differences in returns between urban and rural households; that is, how much rural households would receive if they were endowed with urban characteristics

2010, the contribution of education differences increased across the expenditure distribution from 58 to 63 per cent between the 25th and 75th quantiles. At the median in 2002 (2009/10), differences in education levels accounted for 0.14 (0.09) log points of the differences in per capita expenditure between urban and rural households.

Regional differences were the second largest contributor towards the expenditure gap as an endowment – approximately 30 per cent of the total endowment effect was explained by regional differences. At the median, differences in regional characteristics between urban and rural households contributed by 0.07 log points to the expenditure gap in 2002, and by 0.05 log points in 2009/10. Across the distribution, regional differences contributed more towards the higher welfare gap at the top end of the expenditure distribution in 2002, whereas these differences contributed less at higher quantiles in 2009/10.

Urban-rural differences in endowments by sector of employment (employment in agriculture, manufacturing or services) had the third largest contribution towards the endowment effect in both years. At the median, sectoral differences between urban and rural households caused the expenditure gap to increase by 0.05 log points in 2002 and 0.03 log points in 2009/10. In 2002, the differences by sector of employment contributed by 20 per cent across the distribution, whereas in 2009/10, urban-rural differences in endowments by sector of employment contributed by 13 per cent at the 25th quantile, increasing to a 21 per cent contribution at the 75th quantile. This suggests that such differences have a smaller role to play in explaining the expenditure gap for richer households compared to poorer households.

Differences in the type of employment activity (wage, self-employed, not employed) between urban and rural areas caused the expenditure gap to fall. In 2002 and 2009/10, at the median it explains 6 per cent of the total endowment effect. In 2002 across the distribution, the differences contribute less towards the endowment effect and overall gap at higher quantiles. However in 2009/10, these differences contribute uniformly across the distribution. Other endowments where urban-rural differences in the characteristics contributed towards a fall in the welfare gap include urban-rural differences in age, household size and the proportion of children. On the other hand,

ethnic differences (significant at the 1 per cent level only in 2002) and the transfer of remittance income contribute to a rise in the welfare gap.

Next, some key factors that contributed to the expenditure gap through the differences in *returns* will be discussed. In both years, differences in the returns to employment by age contributed towards a reduction of the "coefficient effect"; although the contribution fell across the expenditure distribution especially in 2002 whereas other factors contributed to a rise in the welfare gap.

Urban-rural differences in returns to education had a large contribution towards the coefficient effect. In 2002, the contribution of the differences in returns to education at the median was 0.38 log points, and in 2009/10 the contribution was 0.17 log points. Across the distribution, the contribution of differences in returns to education fell. In 2009/10, the contribution of this variable at the 25th quantile was 0.16 log points; at the 75th quantile however, differences (0.12 log points) contributed less to a higher endowment effect. This pattern observed by the evolution of the urban-rural gap in returns to education for the lower and middle parts of the expenditure distribution (0.38 log points at the median in 2002 and reducing to 0.17 log points by 2009/10) contribute greatly towards the narrowing of the overall urban-rural gap in per capita expenditure between 2002 and 2009/10.

In 2002, differences in returns to regional characteristics between urban and rural households had a larger impact on richer households, whereas in 2009/10, regional differences had no impact at the 75th quantile, but a large impact on poorer households. This suggests that the differences in returns across regions mattered primarily for the poor in 2009/10. While differences in returns to education and regional characteristics matter less for the richer households in 2009/10, the intercept term is significant at the 1 per cent and contribute greatly towards the welfare gap in 2009/10 at the 75th quantile. This term includes the urban-rural differences in factors not captured in the model. In 2009/10, the contribution of the constant term increased across the expenditure distribution. However in 2002, this constant term was insignificant (at the 10 per cent level) suggesting that the model included covariates that explain the expenditure gap quite accurately. The paper by Thu Le and Booth (2014) noted that the intercept was an important contributor at certain points in the

distribution for the Vietnam analysis; it was stated that factors such as infrastructure and geographic conditions in favour of urban households might be useful in explaining the gap. Unobservable factors, such as those discussed by Thu Le and Booth (2014) can be possible explanations for the contribution of the intercept term towards the overall coefficient effect in 2009/10 in our analysis.

Sectoral differences in returns contribute towards a reduction in the welfare gap, especially in 2009/10. This suggests that the urban-rural differences in returns from working in a particular industry (agriculture, manufacturing and services) were more important in explaining the welfare gap of the rich. Other factors that contributed towards the urban-rural gap include inter-group differences by returns to ethnicity and remittances. The differences in returns by ethnic characteristics explain a substantial part of the overall expenditure gap between urban and rural areas in 2002, however the effects are smaller in 2009/10. Remittance flows contribute by 7 per cent to the total coefficient effect at the 25^{th} quantile, falling to 2 per cent at the 75^{th} quantile in 2002. On the other hand in 2009/10, remittance flows contribute towards a reduction of the coefficient effect by 10 per cent at the highest quantile. Urban-rural differences in returns to demographic characteristics (household size and the presence of children in the household) contribute towards an increase in the welfare gap in 2002, and a decrease in the welfare gap in 2009/10.

The *interaction* effect measures the contribution of simultaneous differences in characteristics and the returns to such characteristics across urban and rural households towards the urban-rural welfare gap. In 2002, the interaction effect increased the welfare gap by 12 per cent at the 25^{th} quantile; however this contribution fell across the distribution – at the 75^{th} quantile, the interaction effect contributed to a fall in the welfare gap by 14 per cent. In 2010, the interaction effect reduced the welfare gap by 7 per cent at the 25^{th} quantile; at the 75^{th} quantile, this effect contributed to a 12 per cent fall in the welfare gap.

To summarize, the decomposition results suggest that the welfare gap rises across the distribution at a given point in time. Over time, the urban-rural welfare gap has fallen. Overall, differences in characteristics such as education, job sector (agriculture, manufacturing and services), regional differences and remittance flows contribute to a

higher welfare gap whereas differences in demographic characteristics (household size and the presence of children) and job type (wage, self-employed, unemployed/not employed) contribute to a lower gap. The differences in *returns* to characteristics changed greatly over time – differences in returns to education, job type and regions contributed less towards the gap at higher quantiles. On the other hand, returns to unobservable characteristics have a large role to play in 2009/10, especially for richer households.

For the purpose of comparison over time, we examined the determinants of per capita expenditure for urban and rural households and estimated the urban-rural welfare gap for the 17 districts included in both years. Tables A5, A6, and A7 in the Appendix include the additional 5 districts and estimate the determinants of urban/rural per capita expenditure and the urban-rural welfare gap for 2009/10[47]. Having examined the determinants of per capita expenditure, the results suggest that the coefficients are lowered when the 5 additional districts are included - the returns to education are lower at any point in the expenditure distribution for both, urban and rural households, compared to the coefficients obtained for the restricted sample. Furthermore, the urban-rural welfare gap is lower – at the median, the gap is 0.261 log points compared to 0.302 log points obtained for the 17 districts. There is a change in the composition of the endowment, coefficient and interaction effects when estimating the welfare gap for all 22 districts in 2009/10. The contribution of the differences in urban-rural characteristics and the contribution of simultaneous differences in urban-rural characteristics and returns to those characteristics towards the total welfare gap is *lower* (43 per cent and -8 per cent respectively at the median, compared to 50 and -12 per cent for the restricted sample). On the other hand, the contribution of the differences in returns to urban-rural characteristics towards the welfare gap is *higher* (65 per cent of the overall gap at the median, in comparison to 62 per cent for the restricted sample).

[47] However, it is important to note that the additional five districts were estimated for certain months of the years.

Table 5: Contributions to the urban-rural expenditure gap at the mean and selected quantiles in 2002 and 2009/10

	2002				2009/10			
	Q25th	Q50th	Q75th	OLS	Q25th	Q50th	Q75th	OLS
Predicted Gap $[\widehat{Y}_u - \widehat{Y}_r]$	**0.423*** (26.55)**	**0.507*** (28.00)**	**0.598*** (30.71)**	**0.500*** (35.66)**	**0.252*** (18.53)**	**0.302*** (21.97)**	**0.333*** (20.69)**	**0.296*** (25.93)**
Due to endowments $[(\widehat{X}_u - \widehat{X}_r)\widehat{\beta}_r]$**:**								
Female	1.78x10^{-4}	-1.88x10^{-4}	-0.001*	-2.57x10^{-4}	-5.98x10^{-5}	-4.82x10^{-4}	-0.001**	-0.001*
	(0.42)	(-0.41)	(-1.90)	(-0.71)	(-0.12)	(-0.95)	(-2.06)	(-1.75)
Sinhalese	0.015***	0.010***	0.006**	0.012***	-2.71x10^{-4}	-3.31x10^{-4}	0.003	0.001
	(5.01)	(3.63)	(2.02)	(5.12)	(-0.14)	(-0.19)	(1.56)	(0.92)
Age	-9.64x10^{-4}	-0.004***	-0.004**	-0.003***	-0.002***	-0.005***	-0.006***	-0.005***
	(-1.31)	(-2.80)	(-2.51)	(-2.67)	(-2.68)	(-3.84)	(-3.46)	(-3.92)
Education	0.096***	0.136***	0.188***	0.148***	0.069***	0.090***	0.117***	0.097***
	(19.24)	(21.63)	(21.75)	(22.99)	(16.16)	(17.68)	(17.62)	(18.50)
Job Type	-0.012***	-0.014***	-0.013***	-0.014***	-0.009***	-0.009***	-0.011***	-0.010***
	(-7.28)	(-7.37)	(-5.59)	(-9.09)	(-6.22)	(-5.75)	(-5.00)	(-7.49)
Job Sector	0.031***	0.050***	0.060***	0.041***	0.015***	0.034***	0.039***	0.027***
	(7.86)	(12.03)	(11.22)	(12.61)	(6.91)	(10.20)	(9.16)	(9.98)
Demographics	-0.023***	-0.031***	-0.035***	-0.031***	-0.015***	-0.018***	-0.020***	-0.018***
	(-6.93)	(-7.29)	(-8.01)	(-7.97)	(-5.39)	(-5.39)	(-5.50)	(-5.50)
Remittances	0.004***	0.005***	0.005***	0.006***	0.004**	0.005***	0.008***	0.006***
	(3.46)	(3.61)	(3.05)	(4.27)	(2.62)	(3.21)	(3.53)	(3.80)
Region	0.045***	0.067***	0.098***	0.067***	0.049***	0.053***	0.060***	0.052***
	(10.29)	(13.32)	(13.48)	(16.35)	(12.73)	(13.30)	(11.34)	(15.67)
Total	**0.155*** (18.41)**	**0.220*** (22.64)**	**0.301*** (24.02)**	**0.225*** (24.67)**	**0.119*** (16.14)**	**0.150*** (19.11)**	**0.187*** (19.13)**	**0.148*** (20.08)**

Due to coefficients $[(\hat{\beta}_u - \hat{\beta}_r)\hat{X}_r]$**:**								
Female	-0.014	0.026	-0.032	-0.026	0.001	-0.031	0.014	-0.012
	(-0.42)	(0.66)	(-0.73)	(-0.97)	(0.03)	(-1.01)	(0.35)	(-0.52)
Sinhalese	0.200***	0.112***	0.018	0.121***	0.043	0.042*	0.046*	0.028
	(6.27)	(3.36)	(0.52)	(5.12)	(1.60)	(1.73)	(1.08)	(1.47)
Age	-0.510***	-0.535***	-0.211*	-0.430***	-0.318***	-0.198***	-0.266***	-0.249***
	(-5.32)	(-5.31)	(-1.81)	(-6.04)	(-4.25)	(-2.67)	(-2.99)	(-4.48)
Education	0.330***	0.384***	0.085*	0.225***	0.162***	0.171***	0.119***	0.121***
	(8.41)	(9.43)	(1.90)	(7.78)	(4.92)	(5.52)	(3.04)	(4.71)
Job Type	0.026	0.125***	0.149***	0.115***	0.034	0.057**	0.034	0.024
	(0.86)	(3.46)	(3.60)	(4.64)	(1.37)	(2.19)	(1.00)	(1.17)
Job Sector	-0.015	-0.110***	-0.086***	-0.071***	-0.054**	-0.086***	-0.122***	-0.067***
	(-0.51)	(-3.62)	(-2.83)	(-3.40)	(-2.30)	(-3.86)	(-4.87)	(-4.04)
Demographics	-0.008	0.052	0.091	0.022	-0.136***	-0.061	-0.037	-0.089***
	(-0.16)	(0.97)	(1.49)	(0.57)	(-3.21)	(-1.47)	(-0.69)	(-2.68)
Remittances	0.016***	0.011**	0.006	0.007**	0.008	$6.48X10^{-4}$	-0.019***	-0.008*
	(3.67)	(2.25)	(1.12)	(2.14)	(1.40)	(0.11)	(-2.70)	(-1.78)
Region	0.040	0.037	0.138**	0.067	0.158***	0.094*	0.078	0.113***
	(0.68)	(0.63)	(2.58)	(1.60)	(2.75)	(1.91)	(1.50)	(2.90)
Constant	0.150	0.195	0.220	0.265**	0.251**	0.198*	0.358***	0.311***
	(1.11)	(1.35)	(1.40)	(2.49)	(2.18)	(1.81)	(2.74)	(3.72)
Total	**0.215*** **	**0.296*** **	**0.379*** **	**0.295*** **	**0.149*** **	**0.188*** **	**0.187*** **	**0.171*** **
	(8.43)	**(11.41)**	**(14.79)**	**(16.62)**	**(7.84)**	**(10.55)**	**(9.49)**	**(12.67)**
Interaction $[(\hat{X}_u - \hat{X}_r)(\hat{\beta}_u - \hat{\beta}_r)]$**:**								
Female	$-4.50x10^{-4}$	$8.24x10^{-4}$	-0.001	$-8.29x10^{-4}$	$-3.46x10^{-5}$	-0.001	$4.81x10^{-4}$	$-4.33x10^{-4}$
	(-0.41)	(0.63)	(-0.69)	(-0.92)	(0.03)	(-0.97)	(0.34)	(-0.50)
Sinhalese	-0.036***	-0.020***	-0.003	-0.022***	-0.006	-0.006*	-0.004	-0.004
	(-5.94)	(-3.32)	(-0.52)	(-4.93)	(-1.60)	(-1.71)	(-1.08)	(-1.46)

Age	-0.008**	-0.006*	$-3.27x10^{-4}$	-0.005	-0.006***	-0.003*	-0.005**	-0.005***
	(-2.62)	(-1.72)	(-0.10)	(-1.69)	(-2.98)	(-1.83)	(-2.37)	(-2.84)
Education	0.079***	0.092***	0.020*	0.054***	0.027***	0.029***	0.020***	0.020***
	(7.99)	(8.84)	(1.89)	(7.44)	(4.77)	(5.32)	(3.00)	(4.58)
Job Type	-0.006	-0.008	-0.013**	-0.011***	-0.008**	-0.008**	-0.006	-0.006**
	(-1.49)	(-1.35)	(-2.01)	(-2.78)	(-2.51)	(-2.20)	(-1.29)	(-2.20)
Job Sector	0.004	-0.048***	-0.048***	-0.029***	-0.015*	-0.033***	-0.047***	-0.023***
	(0.28)	(-3.44)	(-3.36)	(-2.97)	(-1.75)	(-3.84)	(-4.84)	(-3.68)
Demographics	-0.003	-0.010***	-0.004	-0.004*	-0.004***	-0.002	-0.001	-0.003**
	(-1.00)	(-2.82)	(-1.19)	(-1.85)	(-2.77)	(-1.60)	(-0.53)	(-2.21)
Remittances	0.004*	0.001	-0.001	$-4.73x10^{-4}$	0.003	0.002	0.004	0.004*
	(1.79)	(0.48)	(-0.49)	(-0.30)	(0.88)	(0.78)	(1.19)	(1.87)
Region	0.021*	-0.010	-0.032**	-0.003	-0.007	-0.014	-0.002	-0.008
	(1.82)	(-0.84)	(-2.32)	(-0.37)	(-0.71)	(-1.57)	(-0.19)	(-1.11)
Total	**0.053****	**-0.009**	**-0.083*****	**-0.021**	**-0.017**	**-0.036****	**-0.041****	**-0.024****
	(2.57)	**(-0.42)**	**(-3.57)**	**(-1.35)**	**(-1.14)**	**(-2.57)**	**(-2.46)**	**(-2.21)**
Number of observations	16,920	16,920	16,920	16,920	17,182	17,182	17,182	17,182

The decomposition was estimated from the viewpoint of rural households. Base categories: Male (gender), non-Sinhalese (ethnic background), not employed (job type), agriculture (job sector), Sabaragamuwa (region) *z* values in parentheses; Demographic characteristics include household size (log) and the number of children; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; the standard errors are bootstrapped with 500 replications

7 Conclusion

In this chapter, we examined the urban-rural welfare gap in Sri Lanka for the years 2002 and 2009/10. This was a period of transition, falling poverty and significant growth for the economy. The data used for the analysis was from the Household Income and Expenditure surveys carried out in 2002 and 2009/10. The study contributes to existing literature in two significant ways. First, the urban-rural gap has not been formally explored for Sri Lanka. Secondly, the chapter adds to the existing literature on urban-rural welfare gaps through the application of a new method of unconditional quantile regression introduced by Firpo et al. (2009) used to examine the determinants of per capita expenditure for urban and rural households across the distribution of expenditure[48]. Using this method allowed the contributions of individual factors to the overall urban-rural expenditure gap across the distribution to be isolated. In order to do this, a variant of the Blinder-Oaxaca decomposition was applied directly to the results obtained from the unconditional quantile regression technique.

The focus on the entire expenditure distribution instead of looking at the mean was beneficial since the pattern of the gap was not flat across the distribution. The entire distribution was analysed through a quantile regression technique. Across the distribution, the urban-rural welfare gap increased suggesting that there are larger welfare differences across the two areas for richer households in both years.

After identifying the determinants of per capita expenditure for urban and rural households, we observed that both, urban-rural differences in endowments (for example, years of education) and the returns to those endowments (returns to education) contributed towards urban-rural differences in per capita expenditure. The study found that education levels of urban households, which were quite high in the initial period to begin with, remained fairly similar across the analysed period. In the rural sector however, education levels increased. The differences in education levels contributed to a higher welfare gap for richer households in 2009/10. On the other

[48] To the best of my knowledge, this method has been applied to the urban-rural welfare gap in the case of Vietnam. This is based on a literature search in EconLit.

hand regional differences, which also played an important role in affecting the urban-rural welfare gap, contributed less to the welfare gap for richer households in the latter period.

We also found that remittance flows, both local and foreign, have a positive association with per capita expenditure in urban and rural areas; an exception was seen in 2009/10 where local remittances had no significant link with urban per capita expenditure. However, unlike in the case of Vietnam where the receipt of local remittances narrowed the expenditure gap and foreign remittances widened the gap (Thu Le and Booth, 2014); both, differences in foreign and local remittance flows between urban and rural households contributed to a wider welfare gap in the case of Sri Lanka. Other factors that contributed to the expenditure gap include urban-rural differences in ethnicity, household size and age, along with their respective returns.

Between 2002 and 2009/10, the urban-rural welfare gap reduced greatly at every point in the expenditure distribution. Several factors contributed towards this. Differences in certain *endowments* across urban and rural areas reduced during the period. These included regional differences, differences in education levels, job type (wage and self-employment) and job sector (agriculture, manufacturing and services), and differences in demographic characteristics such as household size and the number of children. Differences in *returns* to endowments also narrowed in the latter period. Ethnic differences in returns were no longer significant contributors towards the welfare gap. The differences in returns to education had halved across the distribution by 2009/10; indicating that smaller differences in returns to education contributed towards the overall urban-rural gap in welfare. Another finding in the chapter was that the differences in returns by type of employment (wage and self-employment) were significant in affecting the large urban-rural welfare gap in 2002, but had no significant impact on the gap in 2009/10 for richer households.

Such findings are in line with two important advances in Sri Lanka seen in recent years – industrialization and the government's focus on rural development. The literature indicates that as the country became industrialized, more people started moving away from agriculture to manufacturing and services. This was seen by the rising contribution of services to GDP in recent years and the high returns to working

in services and manufacturing. Our results on the decomposition of the wage gap suggested that, in addition to convergence of the composition of the workforce across urban and rural households between 2002 and 2009/10, the differences in sectoral returns (agriculture, manufacturing and services) across urban and rural areas contributed to a large fall in the welfare gap in the latter year.

By 2009/10, the war had ended and the country prioritized rural development. Infrastructure developments meant that people were able to easily travel to towns for employment. This rural-urban migration in the former years followed up by the end of a 30-year war opening up new opportunities across the country are consistent arguments with the findings of this study where the urban-rural expenditure gap became smaller. During the analysed period, the country's industrialization has been favourable towards rural areas. However, the fact remains that agriculture still gives the lowest returns in employment. Although the significance of this industry has reduced in recent years, it still remains important to a large proportion of rural households. Therefore, if the Sri Lankan government focusses on poverty alleviation and reducing the urban-rural gap, our findings suggest that support to poor rural households that will help them improve agricultural productivity or move into other sectors such as services might be helpful.

We identify certain limitations with the study. Firstly as noted in chapter 2, education and potentially other control variables such as occupation type are endogenous. As this analysis was carried out at the household-level, the individual-level instruments used previously are not appropriate. However as we include household characteristics in this study, we hope that this helps reduce part of the bias caused by disregarding unobservables (for example, by accounting for family background effects directly in the model). As we observed the OLS estimates to be downward biased at the individual level in the previous chapter, we expect the effect of education on the welfare gap to be under-estimated. The second limitation is the lack of information on internal migration and road infrastructure in the HIE surveys. The Institute for Policy Studies (2013) reported that the urban population has remained fairly stable over the years at around 20 per cent. However to the extent that migration took place, it might be a possible explanation for the narrowing of the welfare gap over time. Infrastructure projects may have also contributed to the welfare gap, making it easier

for individuals/produce to be moved across urban and rural areas. Further research to explore the improvements in road infrastructure, migration and movements over time will be useful to uncover these uncertainties. Finally, we note that the end of the war may have had an effect on our results due to the displacement of people in war-affected areas of the country. However, there is no available information on displacement.

Despite the above limitations, this chapter adds to the literature through the examination of the urban-rural welfare gap in Sri Lanka during a period of high growth and falling poverty.

8 References

Blinder, A. S., 1973, "Wage Discrimination: Reduced Form and Structural Estimates", *Journal of Human Resources*, 8 (4), pp. 436–55

Central Bank of Sri Lanka, 2012, Annual Report (link: http://www.cbsl.gov.lk/pics_n_docs/10_pub/_docs/efr/annual_report/AR2012/English/7_Chapter_03.pdf)

Central Bank of Sri Lanka, 2009, Annual Report (link: http://www.cbsl.gov.lk/pics_n_docs/10_pub/_docs/efr/annual_report/ar2009e/PDF/8_chapter_04.pdf)

Deaton, A., 1997, "The Analysis of Household Surveys: A Micro-econometric Approach to Development Policy". Baltimore, MD: Johns Hopkins University Press

Department of Census and Statistics (various statistics including the HIES Final Report, 2009/10)

Firpo, S., Fortin, N.M. and Lemieux, T., 2009, "Unconditional Quantile Regressions," *Econometrica*, 77 (3), pp. 953 – 73

Fortin, N. M., Lemieux, T, and Firpo, S. 2011, "Decomposition Methods in Economics," in O. Ashenfelter, O., Layard, R., and Card, D. (eds), *Handbook of Labor Economics*, Elsevier, Amsterdam, Part A, pp. 1 – 102

Fortin, N. M., 2008, "The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People", *Journal of Human Resources*, 43 (4), pp. 886–920

Himaz, R. and Aturupane, H., 2011, "Education and Household Welfare in Sri Lanka from 1985 to 2006", University of Oxford: Department of Economics Discussion Paper Series, Paper Number: 527

Hnatkovska, V. and Lahiri, A., 2013, "The Rural-Urban Divide in India", Working Paper, International Growth Centre

ILO, "A Strategy for the Rehabilitation of the Rural Transport System in Tsunami-affected areas"
(link: http://www.ilo.org/asia/whatwedo/projects/WCMS_099612/lang--en/index.htm)

Institute for Policy Studies, 2013, "Migration Profile: Sri Lanka"
(link:http://www.ips.lk/ips_migration/publication/migration_profile/migration_profile_ips.pdf)

Knight, J. and Song, L., 1999, "The Rural-Urban Divide: Economic Disparities and Interactions in China", Oxford, Oxford University Press

Koenker, R. and Bassett, G., 1978, "Regression Quantiles", *Econometrica*, 46 (10), pp. 33 – 50

Lanjouw, P. and Ravallion, M., 1995, "Poverty and Household Size", *the Economic Journal*, 105 (433), pp. 1415 – 1434

Lewis, A., 1954, "Economic Development with Unlimited Supplies of Labour*", the Manchester School of Economic and Social Studie*s, 22, pp. 139 – 191

Lewis, A., 1978, "The Evolution of the International Economic Order", Princeton NJ: Princeton University Press

Lipton, M., 1977, "Why Poor People Stay Poor: Urban Bias in World Development", Cambridge MA: Harvard University Press

Lueth, E. and Ruiz-Arranz, M., 2007, "Are Workers' Remittances a Hedge Against Macroeconomic Shocks? The Case of Sri Lanka", IMF Working Paper: WP/07/22

Machado, J. A. F. and Mata, J., 2005, "Counterfactual decomposition of changes in wage distributions using quantile regression", *Journal of Applied Econometrics*, 20 (4), pp. 445 - 465

Meier, G., 1989, "Leading Issues in Economic Development", 5th Edition, New York: Oxford University Press

Mellor, J., 1976, "The New Economics of Growth: A Strategy for India and the Developing World", Ithaca, NY: Cornell University Press

Millennium Development Goals (MDG) Country Report, 2014, United Nations, Sri Lanka

Nguyen, B. T., Albrecht, J. W., Vroman, S. B. and Westbrook, M. D., 2007, "A Quantile Regression Decomposition of Urban–Rural Inequality in Vietnam", *Journal of Development Economics*, 83 (2), pp. 466–90

Oaxaca, R., 1973, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, 14 (3), pp. 693–709

Parker, S., 2009, "The Economics of Self-Employment and Entrepreneurship," Cambridge, Cambridge University Press

Sicular, T., Ximing, Y., Gustafsson, B. and Shi, L., 2007, "The Urban–Rural Income Gap and Inequality in China", *Review of Income and Wealth*, 53 (1), pp. 93–126

Thu Le, H. and Booth, A. L., 2014, "Inequality in Vietnamese Urban–Rural Living Standards, 1993–2006", *Review of Income and Wealth*, 60 (4), pp. 862–886

Uduporuwa, R. J. M, 2010, "An Analysis of Urban Growth and Urbanization in the Sabaragamuwa Province, Sri Lanka", *Sabaragamuwa University Journal*, 9 (1), pp. 115 - 132

United Nations Development Programme (UNDP),
(link: http://www.lk.undp.org/content/srilanka/en/home/countryinfo.html)

World Bank, 2007, "Sri Lanka Poverty Assessment: Engendering Growth with Equity: Opportunities and Challenges", Poverty Reduction and Economic Management Sector Unit: South Asia Region, Report No. 36568-LK

World Bank, 2009, "Sri Lanka Agricultural Commercialization: Improving Farmers' Incomes in the Poorest Regions", Poverty Reduction and Economic Management Sector Unit: South Asia Region, Report No. 48968-LK

Appendix 1: Data on the excluded districts

Five districts were excluded from the analysis as they were not included in the HIES held in 2002 due to the ongoing war. These include Jaffna, Vavuniya, Batticaloa, Ampara and Trincomalee. In order to compare the characteristics of these five districts to those included in the analysis, we explore the within-quintile means for urban and rural households in the excluded districts. The descriptive statistics are presented in Table A1. Some of the key differences are discussed as follows. On average, households in the excluded districts have lower real expenditure per capita at every quantile in comparison to the rest of the country. This is especially true at higher quantiles where large differences in per capita expenditure are observed. The household size in these five districts is smaller, while the proportion of children are higher, compared to the other districts included in the analysis. Adult non-students in these five districts are on average, younger, have lower levels of education, predominantly from a Tamil ethnic background and have not received any form of remittance income. At lower quantiles in urban areas, there are more employed adults in agricultural activities and less in service-related activities compared to the rest of the country – such differences are not present at higher quantiles. In rural areas, the proportion of employed adults in agricultural activities is higher while the proportion of adults in manufacturing and service-related activities is lower, compared to the rest of the country.

Table A1: Within-Quintile Means of Key Variables by log RPCE for urban and rural households in 2009/10 (excluded districts)

2009/10	Urban (1,081 households)					Rural (1,695 households)				
Variable	**1st**	**2nd**	**3rd**	**4th**	**5th**	**1st**	**2nd**	**3rd**	**4th**	**5th**
RPCE (at 2002 prices)	1648	2359	2962	4018	8070	1345	1945	2406	3020	4873
Household size	5.29	4.73	4.59	3.85	3.49	5.26	4.69	4.27	3.98	3.34
Children (proportion)	0.24	0.22	0.22	0.16	0.13	0.27	0.25	0.20	0.18	0.17
Average age	30	31	32	36	40	29	31	33	35	35
Education	4.98	5.90	6.68	7.31	8.74	4.18	4.93	5.31	6.20	6.66
Women (prop.)	0.55	0.55	0.55	0.55	0.55	0.54	0.54	0.52	0.53	0.55
Sinhalese (prop.)	0.02	0.04	0.05	0.06	0.10	0.14	0.12	0.15	0.17	0.18
Local remit (0,1)	0	0	0	0	0	0	0	0	0	0
Foreign remit (0,1)	0	0	0	0	0	0	0	0	0	0
Wage (prop.)	0.20	0.21	0.22	0.23	0.31	0.22	0.22	0.23	0.23	0.25
Self-employed (prop.)	0.07	0.10	0.10	0.13	0.14	0.09	0.10	0.12	0.13	0.14
Not employed (prop.)	0.73	0.69	0.68	0.64	0.55	0.69	0.68	0.65	0.64	0.61
Of the employed:										
Agriculture (prop.)	0.27	0.23	0.18	0.22	0.06	0.63	0.47	0.58	0.40	0.33
Manufacture (prop.)	0.30	0.31	0.26	0.19	0.16	0.19	0.21	0.18	0.15	0.14
Services (prop.)	0.43	0.46	0.56	0.69	0.78	0.28	0.32	0.34	0.45	0.53

The key variable of interest is "real per capita expenditure", calculated in Sri Lankan rupees at 2002 prices (1 USD ≈ 146 LKR). The explanatory variables are computed using adult non-student characteristics. Refer to Table A8 for a detailed explanation of how the variables were constructed. The category of not being in employment is the adult non-students who are unemployed, retired/disabled and stay-at-home parents during the sample period. For the wage and self-employed adults, the sectoral variables capture the proportion of wage/self-employed adults in agriculture, manufacturing or the services sector

Appendix 2

Table A2: District-level Laspeyres price index

District	Price Index		Urban population (%)	
	2002	2009/10	2001	2011
Colombo	1.080	1.071	54.6	77.5
Gampaha	1.060	1.041	14.6	16.3
Kandy	1.020	1.007	12.2	12.1
Galle	1.030	0.993	11.1	12.0
Kalutara	1.070	1.038	10.6	9.2
Puttalama	1.000	1.008	9.2	9.3
Matara	0.980	0.961	8.5	11.8
Matale	0.980	0.985	8.2	14.1
Anuradhapura	0.970	0.980	7.1	5.4
Badulla	0.990	0.990	6.6	8.6
Nuwara-eliya	1.010	1.021	6.1	5.9
Ratnapura	1.020	0.991	5.7	8.9
Hambantota	0.940	0.979	4.1	5.5
Kurunegala	0.950	0.983	2.4	2.0
Kegalle	1.010	1.001	2.2	1.8
Polonnaruwa	0.960	1.001	0	0
Moneragala	0.960	0.974	0	0
Batticaloa		1.065		28.8
Mannar				26.0
Ampara		1.020		23.7
Trincomalee		1.052		22.8
Jaffna		1.075		21.0
Vauniya		1.032		20.7
Mullaitivu				0
Kilinochchi				0

Base for price index: Sri Lanka = 1; the population estimates were obtained from the Census of Population and Housing carried out in 2001 and 2011. The districts were sorted by the proportion of individuals in urban areas in 2001. The Household Income and Expenditure surveys do not disaggregate the districts further to explore urban-rural compositions. The HIES for both years (2002 and 2009/10) were estimated using the census blocks from 2001. The 2011 census has a similar composition, with Colombo being primarily urban and Polonnaruwa and Moneragala being rural. Other districts are fairly similar on the urban-composition scale over the years – exceptions include Matara, Matale, Badulla and Ratnapura which have become more urbanized by 2011. The Northern and Eastern provinces (districts: Mannar, Ampara, Trincomalee, Jaffna, Vavuniya, Mullaitivu and Kilinochchi) were excluded from the census in 2001 due to the war. Therefore, no information on the price index and urban population is available for these areas. By 2009/10, 5 additional districts were included in the survey and by 2011, all districts were included in the census.

Table A3: Quantile regression estimation for the pooled sample at the 25th, 50th and 75th quantiles and the mean

Variables	Q25th		Q50th		Q75th		OLS	
	2002	**2009/10**	**2002**	**2009/10**	**2002**	**2009/10**	**2002**	**2009/10**
Urban (0,1)	0.152***	0.094***	0.230***	0.137***	0.397***	0.196***	0.283***	0.152***
	(12.97)	(8.31)	(16.31)	(11.25)	(17.28)	(11.52)	(25.08)	(15.93)
Personal characteristics (average of non-student adults):								
Female	0.024	0.012	-0.035	-0.037	-0.072*	-0.107***	-0.026	-0.047**
	(1.01)	(0.51)	(-1.35)	(-1.52)	(-1.89)	(-3.35)	(-1.32)	(-2.49)
Sinhalese	-0.070***	0.025	-0.004	0.024*	0.013	-0.023	-0.028**	-0.002
	(-4.38)	(1.63)	(-0.27)	(1.71)	(0.66)	(-1.37)	(-2.25)	(-0.18)
Average age	-0.008***	-0.010***	-0.021***	-0.017***	-0.033***	-0.028***	-0.020***	-0.019***
	(-4.22)	(-5.53)	(-10.00)	(-9.75)	(-11.49)	(-11.98)	(-13.02)	(-13.79)
Average age^2	0.010***	0.010***	0.025***	0.017***	0.042***	0.031***	0.026***	0.020***
	(4.12)	(4.70)	(10.47)	(8.42)	(12.43)	(11.19)	(14.12)	(12.41)
Education	0.067***	0.069***	0.098***	0.088***	0.144***	0.115***	0.105***	0.094***
	(32.99)	(33.62)	(47.15)	(45.48)	(49.30)	(44.95)	(63.00)	(58.65)
Job type (average of non-student adults; default – unemployed/not employed):								
Wage	0.019	0.001	0.031	0.032	0.256***	0.224***	0.145***	0.113***
	(0.79)	(0.04)	(1.22)	(1.32)	(7.03)	(6.85)	(7.31)	(5.80)
Self-employed	0.279***	0.200***	0.270***	0.210***	0.310***	0.245***	0.308***	0.231***
	(9.85)	(7.33)	(8.44)	(7.28)	(6.73)	(6.22)	(12.96)	(10.30)
Job sector (average of non-student adults; default - agriculture):								
Manufacturing	0.053***	0.015	0.057***	-0.029*	-0.060***	-0.089***	0.002	-0.056***
	(2.89)	(0.84)	(3.20)	(-1.79)	(-2.64)	(-4.74)	(0.11)	(-4.44)
Services	0.141***	0.131***	0.191***	0.148***	0.159***	0.134***	0.140***	0.108***
	(9.82)	(9.19)	(13.38)	(11.05)	(8.52)	(8.08)	(12.85)	(10.40)
Other demographic characteristics:								

Children (proportion)	0.045	0.178***	0.166***	0.222***	0.458***	0.316***	0.250***	0.240***
	(1.23)	(5.06)	(4.67)	(6.74)	(9.55)	(7.75)	(9.08)	(9.31)
Household size (log)	-0.370***	-0.382***	-0.462***	-0.444***	-0.522***	-0.499***	-0.449***	-0.444***
	(-30.70)	(-29.86)	(-36.30)	(-36.18)	(-28.04)	(-30.22)	(-44.27)	(-44.42)
Foreign remit (0,1)	0.172***	0.148***	0.211***	0.217***	0.271***	0.324***	0.226***	0.244***
	(7.91)	(7.38)	(8.38)	(10.96)	(6.99)	(11.90)	(11.71)	(15.38)
Local remit (0,1)	0.072***	0.032	0.138***	0.066***	0.160***	0.074***	0.120***	0.051***
	(3.05)	(1.50)	(5.63)	(3.36)	(4.66)	(3.19)	(6.54)	(3.42)
Regions (default – Sabaragamuwa):								
Western	0.234***	0.210***	0.294***	0.248***	0.349***	0.263***	0.258***	0.217***
	(12.40)	(10.34)	(15.81)	(13.36)	(13.93)	(12.11)	(17.91)	(15.02)
Central	0.112***	0.042*	0.107***	0.056***	0.106***	0.101***	0.110***	0.053***
	(5.38)	(1.74)	(5.50)	(2.75)	(4.42)	(4.45)	(7.24)	(3.26)
Southern	0.109***	0.160***	0.124***	0.170***	0.096***	0.167***	0.098***	0.146***
	(4.95)	(7.41)	(5.93)	(8.73)	(3.71)	(7.58)	(6.09)	(9.67)
North West	0.074***	0.071***	0.068***	0.070***	0.044*	0.042*	0.066***	0.042**
	(3.30)	(2.95)	(3.20)	(3.19)	(1.65)	(1.72)	(4.00)	(2.48)
North Central	0.211***	0.144***	0.189***	0.176***	0.123***	0.172***	0.159***	0.140***
	(8.46)	(5.52)	(7.39)	(7.20)	(3.74)	(5.98)	(8.32)	(7.51)
Uva	-0.038	-0.051*	0.033	0.004	0.078***	0.066***	0.039**	-1.10×10^{-4}
	(-1.47)	(-1.78)	(1.45)	(0.18)	(2.77)	(2.61)	(2.15)	(-0.01)
Constant	7.289***	7.588***	7.653***	8.007***	7.929***	8.443***	7.651***	8.072***
	(163.08)	(176.20)	(170.70)	(197.64)	(128.28)	(163.66)	(223.15)	(253.36)
Number of observations	16,920	17,182	16,920	17,182	16,920	17,182	16,920	17,182
R-squared	0.19	0.18	0.29	0.25	0.28	0.24	0.41	0.37

Refer to Table A8 for a detailed explanation of how the variables were constructed; *t* statistics in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; the standard errors are bootstrapped with 500 replications

Table A4: Determinants of per household expenditure per capita at selected quantiles and mean (including interaction terms)

Variables	Q25th		Q50th		Q75th		OLS	
	2002	**2009/10**	**2002**	**2009/10**	**2002**	**2009/10**	**2002**	**2009/10**
Urban (0,1)	0.086	0.111	0.179	0.079	0.317	0.133	0.265***	0.311***
	(0.86)	(1.09)	(1.53)	(0.78)	(1.60)	(0.94)	(2.79)	(3.91)
Personal characteristics (average of non-student adults):								
Female	0.020	-0.016	-0.041	-0.052*	-0.090**	-0.089**	-0.016	-0.041*
	(0.69)	(-0.55)	(-1.37)	(-1.83)	(-2.20)	(-2.47)	(-0.73)	(-1.86)
urban*female	0.021	0.074	0.014	0.054	0.114	-0.045	-0.052	-0.024
	(0.33)	(1.49)	(0.24)	(1.01)	(1.09)	(-0.57)	(-1.02)	(-0.54)
Sinhalese	-0.111***	0.026	-0.050***	0.009	-0.047**	-0.042**	-0.077***	-0.012
	(-5.26)	(1.25)	(-2.66)	(0.51)	(-2.27)	(-2.25)	(-5.24)	(-0.90)
urban*sinhalese	0.131***	0.007	0.159***	0.040	0.162***	0.076**	0.144***	0.036
	(4.39)	(-0.22)	(4.78)	(1.36)	(3.22)	(2.01)	(5.34)	(1.53)
Average age	-0.007***	-0.009***	0.018***	-0.015***	-0.023***	-0.024***	-0.013***	-0.015***
	(-2.95)	(-4.15)	(-7.59)	(-7.52)	(-7.46)	(-9.43)	(7.63)	(-9.31)
urban*age	-0.007*	-0.004	-0.013***	-0.005	-0.052***	-0.011*	-0.035***	-0.015***
	(-1.69)	(-1.05)	(-2.69)	(-1.27)	(-6.38)	(-1.94)	(-8.77)	(-4.71)
Average age^2	0.008***	0.009***	0.022***	0.015***	0.028***	0.027***	0.016***	0.015***
	(2.86)	(3.65)	(7.97)	(6.34)	(7.87)	(8.61)	(8.02)	(8.18)
urban* age^2	0.009*	0.002	0.016***	0.005	0.071***	0.013*	0.048***	0.016***
	(1.88)	(0.53)	(2.79)	(1.15)	(7.48)	(1.87)	(10.18)	(4.35)
Education	0.074***	0.072***	0.099***	0.086***	0.128***	0.105***	0.097***	0.089***
	(31.27)	(28.68)	(42.39)	(37.12)	(39.15)	(35.34)	(52.12)	(46.86)
urban*education	-0.029***	-0.013***	-0.001	0.005	0.069***	0.036***	0.036***	0.018***
	(-6.43)	(-2.98)	(-0.25)	(1.10)	(8.11)	(5.77)	(8.20)	(4.94)

Job type (average of non-student adults; default – unemployed/not employed):								
Wage	0.014	0.035	0.028	0.025	0.186***	0.217***	0.087***	0.109***
	(0.48)	(1.18)	(0.95)	(0.89)	(4.74)	(5.96)	(3.97)	(4.89)
urban*wage	0.045	-0.074	0.005	0.075	0.346***	-0.019	0.282***	0.012
	(0.92)	(-1.43)	(0.08)	(1.31)	(3.13)	(-0.22)	(5.13)	(0.25)
Self-employed	0.287***	0.218***	0.276***	0.206***	0.305***	0.232***	0.294***	0.223***
	(8.78)	(6.71)	(7.64)	(6.17)	(6.21)	(5.30)	(11.38)	(8.82)
urban*Self-employed	-0.090	-0.029	0.002	0.121*	0.293**	0.182*	0.222***	0.129**
	(-1.48)	(-0.51)	(0.02)	(1.77)	(2.03)	(1.73)	(3.20)	(2.27)
Job sector (average of non-student adults; default - agriculture):								
Manufacturing	0.065***	0.018	0.083***	-0.010	-0.035	-0.060***	0.033**	-0.029**
	(3.13)	(0.89)	(4.18)	(-0.58)	(-1.45)	(-3.02)	(2.19)	(-2.09)
urban*manufacturing	-0.099**	-0.074*	-0.155***	-0.137***	-0.129	-0.117**	-0.157***	-0.131***
	(-2.30)	(-1.77)	(-3.00)	(-3.18)	(-1.63)	(-2.04)	(-3.82)	(-4.00)
Services	0.145***	0.137***	0.203***	0.180***	0.204***	0.173***	0.162***	0.131***
	(9.05)	(8.28)	(12.89)	(11.80)	(10.44)	(9.47)	(13.78)	(11.17)
urban*services	-0.092***	-0.081**	-0.105**	-0.177***	-0.199***	-0.138***	-0.107***	-0.107***
	(-2.58)	(-2.38)	(-2.43)	(-4.89)	(-2.89)	(-2.78)	(-3.08)	(-3.86)
Other demographic characteristics:								
Children (proportion)	0.075*	0.202***	0.179***	0.207***	0.400***	0.299***	0.221***	0.237***
	(1.77)	(4.77)	(4.52)	(5.48)	(7.95)	(6.75)	(7.31)	(8.01)
urban*children	-0.102	-0.100	-0.015	0.065	0.359**	0.051	0.203***	-0.004
	(-1.27)	(-1.37)	(-0.17)	(0.83)	(2.51)	(0.48)	(2.84)	(-0.06)
Household size (log)	-0.429***	-0.403***	-0.497***	-0.450***	-0.483***	-0.462***	-0.443***	-0.422***
	(-30.16)	(-26.08)	(-34.17)	(-30.94)	(-23.51)	(-24.50)	(-39.17)	(-36.33)
urban*size	0.277***	0.084***	0.172***	0.042	-0.152***	-0.116***	-0.013	-0.065***
	(11.51)	(3.08)	(5.75)	(1.53)	(-3.08)	(-2.92)	(-0.53)	(-2.86)
Foreign remit (0,1)	0.180***	0.163***	0.184***	0.213***	0.251***	0.307***	0.218***	0.236***

	(6.40)	(6.24)	(5.85)	(8.55)	(5.62)	(9.40)	(9.63)	(12.25)
urban*foreign	-0.054	-0.063	0.072	-0.009	0.048	0.032	0.020	0.005
	(-1.40)	(-1.60)	(1.42)	(-0.22)	(0.54)	(0.55)	(0.46)	(0.14)
Local remit (0,1)	0.064**	0.037	0.124***	0.070***	0.128***	0.076***	0.103***	0.067***
	(2.36)	(1.56)	(4.52)	(3.26)	(3.55)	(3.14)	(5.13)	(4.17)
urban*local	0.047	-0.007	0.073	0.021	0.224**	-0.020	0.120**	-0.095**
	(1.04)	(-0.14)	(1.26)	(-0.39)	(2.24)	(-0.26)	(2.44)	(-2.28)
Regions (default – Sabaragamuwa):								
Western	0.245***	0.228***	0.306***	0.254***	0.372***	0.249***	0.268***	0.221***
	(11.61)	(10.16)	(14.86)	(12.32)	(13.65)	(10.47)	(17.12)	(13.80)
urban*western	-0.151***	-0.009	-0.040	0.043	0.122	0.191***	0.057	0.086**
	(-3.32)	(-0.16)	(-0.67)	(0.80)	(1.41)	(2.77)	(1.26)	(2.04)
Central	0.120***	0.017	0.094***	0.036*	0.071***	0.084***	0.089***	0.039**
	(5.35)	(0.66)	(4.56)	(1.67)	(2.92)	(3.62)	(5.57)	(2.25)
urban*central	-0.135***	0.175***	0.066	0.138**	0.283***	0.147*	0.152***	0.118**
	(-2.65)	(2.69)	(1.02)	(2.20)	(2.95)	(1.83)	(3.03)	(2.42)
Southern	0.128***	0.143***	0.144***	0.155***	0.093***	0.146***	0.109***	0.128***
	(5.32)	(5.95)	(6.40)	(7.25)	(3.52)	(6.30)	(6.39)	(7.80)
urban*southern	-0.180***	0.092	-0.106	0.101*	0.145	0.190***	-0.005	0.130***
	(-3.37)	(1.55)	(-1.61)	(1.79)	(1.49)	(2.65)	(-0.10)	(2.96)
North Western	0.077***	0.054**	0.061***	0.049**	0.025	0.020	-0.056***	0.025
	(3.24)	(2.07)	(2.73)	(2.07)	(0.92)	(0.80)	(-3.27)	(1.41)
urban*north_w	-0.092	0.114*	0.078	0.162**	0.297***	0.251***	0.148**	0.167***
	(-1.58)	(1.68)	(1.06)	(2.48)	(2.59)	(2.99)	(2.57)	(3.22)
North Central	0.228***	0.143***	0.199***	0.172***	0.121***	0.162***	0.165***	0.135***
	(8.49)	(5.09)	(7.35)	(6.63)	(3.57)	(5.41)	(8.30)	(6.83)
urban*north_c	-0.149***	0.031	-0.034	0.056	0.022	0.169	-0.062	0.114*
	(-2.62)	(0.41)	(-0.40)	(0.74)	(0.17)	(1.62)	(-0.91)	(1.87)

Uva	-0.028	-0.072**	0.031	-0.016	0.034	0.040	-0.023	-0.020
	(-1.02)	(-2.39)	(1.31)	(-0.64)	(1.23)	(1.57)	(-1.22)	(-1.02)
urban_uva	-0.090	0.262***	0.025	0.211***	0.393***	0.242**	0.118*	0.215***
	(-1.19)	(3.23)	(0.30)	(2.66)	(3.01)	(2.28)	(1.76)	(3.40)
Constant	7.317***	7.574***	7.678***	8.018***	7.929***	8.418***	7.655***	8.015***
	(143.46)	(150.12)	(153.56)	(173.65)	(121.77)	(148.70)	(204.69)	(221.74)
Number of observations	16,920	17,182	16,920	17,182	16,920	17,182	16,920	17,182
R-squared	0.20	0.19	0.29	0.26	0.29	0.24	0.42	0.37

Refer to Table A8 for a detailed explanation of how the variables were constructed; *t* statistics in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; the standard errors are bootstrapped with 500 replications

Table A5: Determinants of urban household expenditure per capita at the mean and selected quantiles in 2009/10 (all districts)

Variables	Q25th	Q50th	Q75th	OLS
Personal characteristics (average of non-student adults):				
Female	0.005	-0.068	-0.072	-0.067*
	(0.12)	(-1.44)	(-1.23)	(-1.91)
Sinhalese	0.060**	0.063**	0.065**	0.056***
	(2.47)	(2.48)	(2.22)	(2.93)
Average age	-0.016***	-0.020***	-0.030***	-0.024***
	(-5.00)	(-5.65)	(-6.95)	(-9.34)
Average age^2	0.016***	0.022***	0.034***	0.026***
	(4.31)	(5.41)	(6.61)	(8.59)
Education	0.077***	0.098***	0.112***	0.097***
	(20.87)	(25.98)	(24.21)	(32.62)
Job type (average of non-student adults; default – unemployed/not employed):				
Wage	-0.012	0.111**	0.196***	0.086**
	(-0.26)	(2.11)	(2.87)	(2.13)
Self-employed	0.254***	0.330***	0.372***	0.298***
	(4.88)	(5.22)	(4.57)	(6.24)
Job sector (average of non-student adults; default - agriculture):				
Manufacturing	-0.035	-0.011	-0.050	-0.035
	(-0.95)	(-1.00)	(-1.21)	(-1.15)
Services	0.116***	0.141***	0.168***	0.143***
	(3.77)	(3.58)	(4.53)	(4.03)
Other demographic characteristics:				
Children (proportion)	0.201***	0.217***	0.299***	0.200***
	(3.08)	(3.69)	(3.96)	(4.10)
Household size (log)	-0.435***	-0.486***	-0.475***	-0.474***

	(-19.31)	(-20.55)	(-16.07)	(-25.73)
Foreign remit (0,1)	0.226***	0.270***	0.246***	0.259***
	(6.86)	(7.24)	(5.22)	(8.93)
Local remit (0,1)	0.069	0.016	-0.089	-0.021
	(1.34)	(0.27)	(-1.38)	(-0.52)
Regions (default – Sabaragamuwa):				
Western	0.277***	0.295***	0.368***	0.304***
	(4.82)	(5.60)	(6.25)	(7.39)
Central	0.204***	0.143**	0.193***	0.167***
	(3.10)	(2.28)	(2.80)	(3.49)
Southern	0.278***	0.231***	0.302***	0.262***
	(4.70)	(4.18)	(4.87)	(6.13)
Northern	0.250***	0.162**	0.230***	0.224***
	(3.52)	(2.37)	(3.02)	(4.39)
Eastern	0.169**	0.045	0.151**	0.132***
	(2.60)	(0.78)	(2.36)	(2.88)
North West	0.216***	0.197***	0.200***	0.203***
	(3.13)	(3.05)	(2.75)	(3.96)
North Central	0.166**	0.208***	0.327***	0.247***
	(2.11)	(2.69)	(3.43)	(4.08)
Uva	0.174**	0.184**	0.172*	0.199***
	(2.06)	(2.22)	(1.89)	(3.14)
Constant	7.742***	8.132***	8.522***	8.240***
	(84.70)	(88.75)	(78.48)	(119.85)
Number of observations	5,273	5,273	5,273	5,273
R-squared	0.19	0.26	0.23	0.36

Refer to Table A8 for a detailed explanation of how the variables were constructed; *t* statistics in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; the standard errors are bootstrapped with 500 replications.

Table A6: Determinants of rural household expenditure per capita at the mean and selected quantiles in 2009/10 (all districts)

Variables	Q25th	Q50th	Q75th	OLS
Personal characteristics (average of non-student adults):				
Female	-0.024	-0.024	-0.075**	-0.043**
	(-0.97)	(-0.99)	(-2.34)	(-2.22)
Sinhalese	0.013	-0.016	-0.002	0.004
	(0.75)	(-1.03)	(-0.12)	(0.35)
Average age	-0.004**	-0.013***	-0.022***	-0.013***
	(-2.38)	(-7.21)	(-9.44)	(-9.16)
Average age^2	0.003	0.012***	0.024***	0.013***
	(1.60)	(5.84)	(8.55)	(7.77)
Education	0.060***	0.077***	0.101***	0.084***
	(28.58)	(38.25)	(38.15)	(49.53)
Job type (average of non-student adults; default – unemployed/not employed):				
Wage	0.014	0.017	0.146***	0.092***
	(0.54)	(0.69)	(4.38)	(4.50)
Self-employed	0.159***	0.175***	0.224***	0.201***
	(5.86)	(6.10)	(5.65)	(8.77)
Job sector (average of non-student adults; default - agriculture):				
Manufacturing	0.010	-0.001	-0.060***	-0.033**
	(0.57)	(-0.08)	(-3.23)	(-2.60)
Services	0.108***	0.156***	0.174***	0.123***
	(7.81)	(11.75)	(10.37)	(11.75)
Other demographic characteristics:				
Children (proportion)	0.148***	0.169***	0.272***	0.216***
	(4.15)	(5.14)	(6.70)	(8.13)
Household size (log)	-0.360***	-0.422***	-0.475***	-0.420***
	(-27.81)	(-33.93)	(-28.50)	(-40.45)

Foreign remit (0,1)	0.135***	0.197***	0.288***	0.234***
	(5.86)	(8.56)	(9.16)	(12.69)
Local remit (0,1)	0.027	0.052***	0.065***	0.062***
	(1.24)	(2.67)	(2.80)	(4.05)
Regions (default – Sabaragamuwa):				
Western	0.184***	0.224***	0.293***	0.225***
	(9.29)	(11.64)	(12.30)	(14.72)
Central	0.008	0.027	0.087***	0.040**
	(0.36)	(1.29)	(3.63)	(2.36)
Southern	0.112***	0.142***	0.154***	0.126***
	(5.24)	(7.02)	(6.50)	(7.98)
Northern	0.064	-0.004	-0.046	0.006
	(1.57)	(-0.11)	(-1.10)	(0.19)
Eastern	0.048*	0.012	-0.007	0.008
	(1.77)	(0.52)	(-0.28)	(0.41)
North West	0.049**	0.045**	0.032	0.025
	(2.09)	(2.05)	(1.21)	(1.45)
North Central	0.122***	0.154***	0.176***	0.134***
	(4.92)	(6.36)	(5.81)	(7.05)
Uva	-0.058**	-0.022	0.018	-0.024
	(-2.11)	(-0.93)	(0.69)	(-1.26)
Constant	7.610***	8.021***	8.396***	8.025***
	(176.15)	(195.90)	(160.60)	(243.54)
Number of observations	14,685	14,685	14,685	14,685
R-squared	0.16	0.23	0.22	0.34

Refer to Table A8 for a detailed explanation of how the variables were constructed; *t* statistics in parentheses; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$; the standard errors are bootstrapped with 500 replications

Table A7: Contributions to the urban-rural expenditure gap at the mean and selected quantiles in 2009/10 (all districts)

	Q25th	Q50th	Q75th	OLS
Predicted Gap	**0.224*** **	**0.261*** **	**0.314*** **	**0.271*** **
$[\widehat{Y}_u - \widehat{Y}_r]$	**(19.46)**	**(21.09)**	**(21.78)**	**(26.86)**
Due to endowments $[(\widehat{X}_u - \widehat{X}_r)\widehat{\beta}_r]$**:**				
Female	-4.58x10^{-4}	-4.56x10^{-4}	-0.001**	-0.001**
	(-0.94)	(-0.96)	(-2.14)	(-2.04)
Sinhalese	-0.002	-0.003	3.28x10^{-4}	-0.001
	(-0.75)	(-1.03)	(0.12)	(-0.35)
Age	-0.001**	-0.004***	-0.005***	-0.004***
	(-2.54)	(-3.48)	(-3.19)	(-3.51)
Education	0.061***	0.079***	0.104***	0.086***
	(16.82)	(18.27)	(18.26)	(19.18)
Job Type	-0.007***	-0.008***	-0.009***	-0.009***
	(-5.47)	(-5.65)	(-5.13)	(-7.45)
Job Sector	0.021***	0.030***	0.033***	0.023***
	(7.23)	(10.34)	(8.93)	(9.95)
Demographics	-0.014***	-0.017***	-0.019***	-0.017***
	(-5.57)	(-5.60)	(-5.72)	(-5.73)
Remittances	0.002*	0.003**	0.004**	0.003**
	(1.82)	(1.99)	(2.42)	(2.41)
Region	0.033***	0.032***	0.036***	0.034***
	(9.40)	(8.82)	(7.78)	(11.12)
Total	**0.092*** **	**0.113*** **	**0.142*** **	**0.116*** **
	(14.01)	**(15.90)**	**(16.23)**	**(17.32)**
Due to coefficients $[(\widehat{\beta}_u - \widehat{\beta}_r)\widehat{X}_r]$**:**				
Female	0.015	-0.023	0.002	-0.012

	(0.59)	(-0.83)	(0.05)	(-0.59)
Sinhalese	0.034	0.033	0.048**	0.037**
	(1.60)	(1.58)	(1.99)	(2.28)
Age	-0.207***	-0.087	-0.108	-0.169***
	(-3.28)	(-1.32)	(-1.38)	(-3.48)
Education	0.113***	0.130***	0.070**	0.082***
	(4.17)	(4.75)	(2.05)	(3.75)
Job Type	0.008	0.048**	0.036	0.013
	(0.39)	(2.09)	(1.20)	(0.76)
Job Sector	-0.006	-0.059***	-0.081***	-0.033**
	(-0.31)	(-3.27)	(-3.94)	(-2.43)
Demographics	-0.093**	-0.073*	0.004	-0.076**
	(-2.60)	(-1.90)	(0.08)	(-2.59)
Remittances	0.008	0.001	-0.014**	-0.005
	(1.64)	(0.20)	(-2.25)	(-1.32)
Region	0.129**	0.089*	0.118**	0.119***
	(2.39)	(1.81)	(2.16)	(3.08)
Constant	0.132	0.110	0.126	0.214***
	(1.30)	(1.10)	(1.04)	(2.81)
Total	**0.133*** **	**0.170*** **	**0.201*** **	**0.170*** **
	(8.33)	**(10.86)**	**(11.30)**	**(14.41)**
Interaction $[(\widehat{\boldsymbol{X}}_{\boldsymbol{u}} - \widehat{\boldsymbol{X}}_{\boldsymbol{r}})(\widehat{\boldsymbol{\beta}}_{\boldsymbol{u}} - \widehat{\boldsymbol{\beta}}_{\boldsymbol{r}})]$**:**				
Female	$-5.59\text{x}10^{-4}$	-0.001	$6.03\text{x}10^{-5}$	$-4.56\text{x}10^{-4}$
	(0.58)	(-0.80)	(0.05)	(-0.58)
Sinhalese	-0.008	-0.007	-0.011*	-0.008**
	(-1.60)	(-1.58)	(-1.90)	(-2.27)
Age	-0.004**	-0.001	-0.001	-0.002**
	(-2.28)	(-0.52)	(-0.79)	(-2.03)

Education	0.018***	0.021***	0.011**	0.013***
	(4.08)	(4.63)	(2.04)	(3.69)
Job Type	-0.004*	-0.006**	-0.006	-0.004*
	(-1.68)	(-2.17)	(-1.60)	(-1.84)
Job Sector	$8.27x10^{-4}$	-0.020***	-0.030***	-0.011**
	(0.12)	(-3.01)	(-3.87)	(-2.19)
Demographics	-0.003**	-0.003**	$-1.10x10^{-4}$	-0.002**
	(-2.54)	(-2.24)	(-0.08)	(-2.09)
Remittances	$4.29x10^{-4}$	0.003	0.005*	0.004**
	(0.17)	(1.16)	(1.72)	(2.06)
Region	-0.003	-0.007	0.002	-0.004
	(-0.32)	(-0.86)	(0.25)	(-0.68)
Total	**$-4.45x10^{-4}$**	**-0.021***	**-0.029****	**-0.015**
	(-0.04)	**(-1.73)**	**(-1.99)**	**(-1.61)**
Number of observations	19,958	19,958	19,958	19,958

The decomposition was estimated from the viewpoint of rural households. Base categories: Male (gender), non-Sinhalese (ethnic background), not employed (job type), agriculture (job sector), Sabaragamuwa (region) *z* values in parentheses; Demographic characteristics include household size (log) and the number of children; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; the standard errors are bootstrapped with 500 replications

Table A8: Variables used for the main analysis: HIES 2002 and 2009/10

Variable	Description	Number of observations	Mean	Minimum	Maximum	Standard deviation
HIES 2002						
real per capita expenditure	monthly per capita expenditure by the household	16,920	3,284	274	108,633	3,925
household size	number of persons in the household	16,920	4.36	1	19	1.77
proportion of children (age<15)	the proportion of children in the household below the age of 15	16,920	0.19	0	0.83	0.20
urban	urban=1 if the household is located in an urban area, else zero (rural)	16,920	0.19	0	1	0.39
local remittance	=1 if household receives remittance income within the country, else zero	16,920	0.05	0	1	0.22
foreign remittance	=1 if household receives remittance income from abroad, else zero	16,920	0.05	0	1	0.21
Characteristics of adult non-students in the household:						
education (years)	average years of education for adult non-students	16,920	6.63	0	19	3.06
age	average age for adult non-students	16,920	34.68	15	65	13.42
female	proportion of female adult non-students	16,920	0.50	0	1	0.21
Sinhalese	proportion of Sinhalese adult non-students	16,920	0.81	0	1	0.36
primary industry	proportion of adults working in primary industries	16,920	0.29	0	1	0.42
secondary industry	proportion of adults working in secondary industries	16,920	0.18	0	1	0.35
tertiary industry	proportion of adults working in tertiary industries	16,920	0.44	0	1	0.46
wage employment	proportion of adults working in wage employment	16,920	0.29	0	1	0.27
self-employment	proportion of adults working in self-employment	16,920	0.14	0	1	0.21
unemployed/not	proportion of adults who are unemployed/not	16,920	0.57	0	1	0.26

employed	employed					
HIES 2009/10						
real per capita expenditure	monthly per capita expenditure by the household (at 2002 prices)	17,182	4,090	396	347,855	5,652
household size	number of persons in the household	17,182	4.27	1	17	1.73
proportion of children (age<13)	the proportion of children in the household below the age of 15	17,182	0.17	0	0.75	0.19
urban	urban=1 if the household is located in an urban area, else zero (rural)	17,182	0.24	0	1	0.43
local remittance	=1 if household receives remittance income within the country, else zero	17,182	0.08	0	1	0.26
foreign remittance	=1 if household receives remittance income from abroad, else zero	17,182	0.07	0	1	0.25
Characteristics of adult non-students in the household:						
education (years)	average years of education for adult non-students	17,182	6.80	0	19	3.08
age	average age for adult non-students	17,182	36.21	15	65	14.67
female	proportion of women adult non-students	17,182	0.52	0	1	0.21
Sinhalese	proportion of Sinhalese adult non-students	17,182	0.75	0	1	0.39
primary industry	proportion of adults working in primary industries	17,182	0.26	0	1	0.41
secondary industry	proportion of adults working in secondary industries	17,182	0.20	0	1	0.36
tertiary industry	proportion of adults working in tertiary industries	17,182	0.43	0	1	0.46
wage employment	proportion of adults working in wage employment	17,182	0.26	0	1	0.27
self-employment	proportion of adults working in self-employment	17,182	0.15	0	1	0.21
unemployed/not employed	proportion of adults who are unemployed/not employed	17,182	0.59	0	1	0.26

Chapter 4: Sectoral choice and sample selectivity

1 Introduction

This chapter examines the characteristics and earnings of individuals employed in different sectors in the context of Sri Lanka. Men and women are analysed separately as labour force participation decisions and earnings are expected to differ between the sexes. Additionally, the chapter deals with two forms of possible biases which have not been examined simultaneously in the context of Sri Lanka in previous literature[49] – (1) the possibility of an endogenous sample selectivity bias regarding labour force participation and sector choice, and (2) the possible endogeneity of education in the earnings functions.

This chapter is motivated by findings from the previous chapters. Having analyzed the returns to education in chapter 2, it was clear that individuals' returns to education varied across different types of employment and across the sexes. The analysis on the falling urban-rural welfare gap between 2002 and 2009/10 suggested that education and employment choice were two of the main factors contributing towards the welfare gap, and smaller urban-rural differences in these two factors, among others, gave rise to a smaller urban-rural welfare gap in the latter period of analysis. Such findings on the outcomes of the labour market motivate the analysis of this chapter – where we identify the factors that drive labour force participation for men and women, and how this varies across different sectors such as private formal and informal sectors, public sector, agriculture and self-employment.

Standard human capital theory suggests that wage differentials arise from differences in human capital endowments across individuals. This has implications for the labour market, especially in developing countries where the labour market is segmented. Four main categories have been identified – rural/agriculture, public, private formal and informal - differentiated by the types of job contracts, structure of earnings, seasonality of production, and uncertainty of demand. Research on developed

[49] This was revealed from a literature search in EconLit

countries traditionally focusses on formal sector employment when examining wage differentials. Bennell (1996) states that focusing on the formal sector employees while ignoring the rural and informal sector employees' in developing countries can lead to an imprecise representation of the labour market. Furthermore, self-employment is less likely to be linked to individual qualities or to a pay scale; Vijverberg (1995) observed that education plays a minor role in explaining the earnings of such individuals. If the returns differ across these sectors/segments, perhaps individual characteristics such as education can help or deter a person from getting in through the entry barriers into a high-paying sector. Consequently, it would be wrong to overlook the existence of differences across segments of the labour market when identifying the link between education and labour market participation and outcomes. The research question for this chapter stems from this area of the literature.

The chapter contributes to the existing literature by controlling for a possible sample selectivity bias when estimating earnings. Gunatilaka (2008) estimated the earnings functions while controlling for sample selectivity in Sri Lanka during 2006. However the paper included only wage employees, and excluded the self-employed. Additionally, education was treated as exogenous. As noted in chapter 2, ignoring the endogeneity of education leads to biased OLS estimates. Thus, we add to the existing literature by dealing with both forms of bias and including all individuals in the labour force.

The aim of this chapter is threefold. First, to analyse the determinants of sector choice and earnings across sectors we address the issues of sample selectivity and endogeneity of education. Secondly, in order to look at differences across sectors, we disaggregate the labour market into five sectors of employment - non-agriculture sectors include the public sector, wage employment in the private formal and informal sectors, self-employment in the private sector, and work related to agricultural activities (wage, self-employment, private formal and informal) form the final sector. Thirdly, as we expect differences in determinants of sector choice and earnings across men and women, this chapter examines the two sexes separately.

The data used in this study comes from the Sri Lankan Labour Force Survey 2013. This data set is the first in its series to include detailed information on income for the

self-employed thus allowing us to look at the informal sector in greater detail. In order to deal with the potential issue of sample selectivity while examining the determinants of sector choice, Lee's multinomial logit (MNL) estimation is employed. The MNL allows us to understand the individual characteristics that determine which sector an individual would join into. This part of the analysis includes individuals who are not in the labour force in addition to those who are. By doing so, we are able to obtain predicted probabilities that are included in the earnings functions in the second part of the analysis when the earnings functions are estimated. Hence we control for the potential bias of sample selectivity.

To estimate the earnings across sectors, Mincerian earnings functions are estimated using Ordinary Least Squares (OLS). The earnings functions control for the possible biases of sample selectivity and endogeneity of education. In order to deal with the latter issue, a Control Function (CF) approach is used – education is regressed on all exogenous variables, the residuals are obtained and then included in the MNL estimation of labour force participation and OLS estimation of the earnings functions. This approach has been used in previous studies by Garen (1984), Wooldridge (2005), Söderbom et al. (2006) and Kuépié et al. (2009).

The findings suggest that endogeneity of education is a greater issue than sample selectivity in the context of Sri Lanka. We also find that more educated individuals get into the public and formal private sectors. Individuals who choose employment in agricultural or the informal sector have the lowest levels of education. Married men are more likely to join the labour force, whereas married women are less likely to join. The earnings functions suggest that the rates of returns to education vary greatly across sectors – an extra year of education yields the highest earnings in the formal private sector and the lowest in the informal and agricultural sectors. The earnings, conditional on observable characteristics, suggest that the public sector pays more to women than men on average; however in the other sectors, men are paid more than women on average. As the characteristics and earnings of individuals vary across the five sectors, this validates the need to explore sectors in isolation. Further, as men and women differ in their decision to participate in the labour force and also the returns to employment, analysing them separately is important.

The rest of the chapter is structured as follows. Section 2 reviews the existing literature. Section 3 describes the data, and Section 4 specifies the methods. Section 5 presents the main findings and Section 6 concludes.

2 Literature review

This section summarizes previous research on the wage differentials across sectors and the determinants of sector choice. Given the importance of method in this literature, the review focusses on econometric issues raised by the various studies that are subsequently used to guide the empirics of this chapter.

2.1 Wage premium in the public sector

In many developing, as well as developed countries, a public sector wage premium has been observed. Several studies have identified this premium for countries such as Pakistan, Greece, Poland and Canada (Hyder and Reilly, 2005; Christopoulou and Monastiriotis, 2014; Adamchik and Bedi, 2000; Tiagi, 2010). A widely used explanation for the premium has been that wages are politically determined in the public sector whereas they are based on productivity and profits in the private sector (Gunderson, 1979). Political forces indirectly influence the setting of public sector wages via institutional channels. Decisions to limit the growth and spending of the public sector, intergovernmental transfers and other policies can also have an influence on wages in the public sector. Another explanation is that trade unions possibly exhibit greater freedom in the public sector since services provided by this sector are essential and thus, labour demand is inelastic (Heitmueller, 2004). The substantial provision of non-wage benefits, job security and the higher degree of unionization have been identified as reasons for the preference towards public sector jobs (Mengistae, 1999). The literature suggests that people prefer jobs in the public sector, and would not mind queuing up for such jobs.

Fogel and Lewin (1974) argued that wages differ across the two sectors due to the implications of the difficulty in assessing the relative worth of the public sector since many of its activities are not marketable. To resolve this issue, wage setters in the

public sector base the wages on the wage given for comparable activities in the private sector. However, because the private sector may not operate in a perfectly competitive setting, this leads to a range of wage rates above or below the wages that would prevail in a perfectly competitive market. The wage setters in the public sector are then faced with a menu of wage rates to choose from. Fogel and Lewin suggest that this responsibility to choose from a range of wage rates, along with the political processes involved in wage setting leads to public sector wages that are higher on average than private sector wages.

2.1.1 Methods

A popular empirical framework used to examine the public sector wage premium is the endogenous switching regression. Adamchik and Bedi (2000), Christopoulou and Monastiriotis (2014) and Tiagi (2010) employed this model for Poland, Greece and Canada respectively. This framework uses a reduced-form probit model to identify the characteristics of individuals who join into the public and private sectors, known as the selection equation. The wage functions are estimated thereafter, for each sector. The selection equation is useful to understand whether more/less educated individuals, men or women, older or younger individuals are likely to join the public or private sectors. Controlling for selection in the wage equations, it is then possible to identify whether self-selection is rewarded or penalized in each sector in terms of earnings. In the study on Greece, Christopoulou and Monastiriotis (2014) found that men in the public sector are greatly rewarded compared to male counterparts in the private sector – the latter group received lower returns compared to women as well; women earned fairly similar returns across sectors. The study was able to identify a great degree of gender discrimination in the public sector - women received an education premium that is 30 per cent lower than the premium that men receive. However, such differences were not found in the private sector. In Canada, Tiagi (2010) observed that individuals who join the public sector are positively selected into that sector, thus earn more, whereas individuals who join the private sector are negatively selected – this was true for both, men and women; however the size of the public-sector premium was higher for women in comparison to men.

The switching regression framework requires at least one exclusion restriction, that is, at least one variable that influences sector choice but has no influence on wages. Adamchik and Bedi (2000) used age and entry into the labour market after the year 1989 as variables that influence sector choice but not wages – it was believed that post-1989 entrants into the labour market are more likely to work in the private sector. However it is not clear as to why age should not influence wages, when there is a vast literature on the effects of age on earnings. Christopoulou and Monastiriotis (2014) and Tiagi (2010) used household size and dummy variables to identify whether the individual has more than one job and whether he/she has a parent or spouse working in, or retired from, the public sector as exclusion restrictions in their analysis of the switching regression in Greece and Canada. The difficulty in empirically testing whether the chosen variables are valid exclusion restrictions is a limitation in the switching regression framework.

Studies that ignore the non-participants and only control for the sample selection across sectors can still lead to biased estimates. Heitmueller (2004) employed a double-selection model where sample selection is controlled for in the decision to participate in the labour force *and* in the choice of sector (public or private). The additional controls were included in the wage functions for each sector. In order to achieve identification in the labour force participation decision, this paper included controls for the number of children in two age categories (0-11, 12-18). These controls were excluded from the sector choice and wage equations. To achieve identification in the sector choice equation, controls for union status were included; these controls were excluded from the participation and wage equations. Heitmueller (2004) found that there was no sample selection from the decision to participate in the labour force for both men and women. Sample selection from the choice of sector had significant effects on the wage functions, especially for men in the public and private sectors. Terrell (1993) employed a double-selection model where an individual first chooses either the public or private sector, and individuals who join the public sector then choose between the public administration and state-owned enterprise employment in Haiti. However, it is less clear as to why this is sequential process; it is possible that workers join the public sector and the type of occupation within the public sector simultaneously, rather than sequentially. Therefore, we estimate Mincerian earnings functions that exclude occupation choice as we believe that

occupation and sectoral choice are more likely to occur simultaneously. This is discussed in greater detail in subsequent sections.

2.2 The role of the informal sector

The studies discussed so far look at two sectoral choices – public or private. However, it is well-known that the labour market in developing countries is disaggregated into agriculture/rural, public, private formal and private informal sectors since each of these sectors have unique characteristics such as job uncertainty and seasonality about demand, the structure of wages/earnings and the nature of contracts (Hess and Ross, 1997; Ray, 1998; Schultz, 2004). Gunatilaka (2008) identified that the characteristics of workers employed in the formal sector differ from the characteristics of informal sector workers in Sri Lanka. Vijverberg (1995) observed the minor role that education plays in explaining earnings of self-employed individuals as it is believed that such employment is less likely to be linked to individual qualities or to a pay scale.

The importance of informal employment has been identified in numerous studies. This sector was conventionally associated with poor quality and lower wages. More recently, there has been evidence of the growth of incomes among those working in this sector in developing countries. Dasgupta (2003) studied the informal service employment in New Delhi, in which she found that the average earnings of informal service sector workers are not the lowest in the economy. Pratap and Quintin (2006) observed a formal sector wage premium in Argentina. After controlling for individual and employer characteristics, a standard wage regression suggested that the premium still existed. However, the premium disappeared when semi-parametric techniques based on propensity score matching were used, where formal sector workers and informal sector workers with similar propensity scores were matched. In several sub-samples, the paper found that informal workers earn more than their formal counterparts.

2.2.1 Methods

Studies discussed previously in this section used a probit model to look at selection into the public or private sectors, whereas Tansel (2004) and Gunatilaka (2008) used a multinomial logit model which allowed them to further disaggregate the private sector – in Turkey and Sri Lanka, respectively. Additionally, this method allows for the inclusion of non-participants in the labour force, thus accounting for sample selectivity bias from both, the sectoral choice *and* the decision to participate in the work force. The study by Gunatilaka (2008) was able to distinguish between the formal and informal sectors in Sri Lanka. The paper focussed on wage earners (due to a lack of data on the self-employed). It first analysed the probability of employment in the public sector, formal and informal private sectors, and the determination of wages in each sector after controlling for the selectivity bias. The paper included several variables that were thought to affect the probability of employment, but not the earnings – namely, the number of employed members of the household, marital status, the presence of children and elderly in the households. The study found that individuals with less education, in agricultural employment and men are more likely to be informally employed. The findings on the determinants of wages include higher earnings for men, more educated individuals, and those employed in manufacturing and services industries (particularly in the formal sector).

Within this and the previous literature, selectivity bias has been taken into account - that is, the non-random way in which individuals get into various sectors of employment - when estimating the wage functions for each sector. Hyder and Reilly (2005) estimated the wage functions for the public and private sectors and state-owned enterprises in Pakistan after correcting for the selectivity bias. In order to account for this form of bias, Lee's (1983) multinomial logit model was employed. The public sector was found to have more educated individuals and a smaller gender pay gap compared to the private sector. A further finding from this study was that there was positive selection into the private sector, that is, individuals who joined the private sector earned higher wages, but no significant selection effects were observed in the public sector.

Variations of the Heckman selection model have been used in the studies mentioned above. These applications implicitly assume that the covariates in the model were exogenous. However, as we observed in chapter 2, this is a questionable assumption. It is possible for unobservable characteristics such as ability and family background to affect the probability of labour force participation, education and (potential) earnings jointly – thus, education cannot be regarded as exogenous in such a case. More recent studies (for example, Kuepie et al., 2009; Schwiebert, 2005) analysed the determinants of labour force participation and earnings while accounting for possible endogeneity of education. Schwiebert (2015) analysed the female workforce in the U.S in the late 1970s using Heckman's (1979) model with the inclusion of an endogenous covariate (education). The findings of the paper suggested that endogeneity of education led to substantially higher returns to education for the female employees; compared to a similar study by Mulligan and Rubinstein (2008) using the same data which assumed education to be exogenous. However, Schwiebert (2015) found that Mulligan and Rubinstein's (2008) finding that the female workforce was negatively selected is robust to accounting for potential endogeneity of education.

Kuepie et al. (2009) observed an increase in the returns to education once endogeneity was accounted for, particularly in the informal sector. The paper used a Control Function (CF) approach, rather than instrumental variables (IV). Variables such as fathers' level of schooling and main occupation were included as instruments to estimate education; this was used as a control for potential endogeneity of education. The endogenous education was then included in the selection model and earnings functions. Studies by Garen (1984), Söderbom et al. (2006) and Wooldridge (2005) employed the CF technique. Additionally, Kuepie et al. (2009) who employed this technique for several West African cities observed that the assumption that education is exogenous is rejected in most cases, apart from the public sector in certain cities. Further, the paper provided evidence of more refined estimates of the returns to education in all cities and sectors after correcting for selectivity bias.

The papers mentioned above, while controlling for possible endogeneity and sample selection biases, used different variables to address the two issues – that is, different sets of instruments to account for the two forms of bias. Kuepie et al. (2009), for instance, estimated education on father's education level and occupation as

instruments along with individual (exogenous) characteristics. To control for selectivity, they incorporated dummy variables which identified the individual's relationship with the head of the household and the household's inverse dependency ratio along with exogenous characteristics and education (controlling for potential endogeneity). However as Wooldridge (2010) discusses, in this way we are simply choosing the variables that are viewed as instruments for education and those that affect selection. By doing so, if the exclusion restrictions made in the selection model are violated, the wage estimation can be inconsistent. To avoid this issue, Wooldridge suggests including all variables in the estimation of education and selection when accounting for both forms of bias.

2.3 Differences across men and women

The acknowledgement of gender differences in wages has led to a number of studies which examine the earnings and determination of labour force participation for men and women separately. It is common to see the use of the Blinder-Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973) to explore the gender wage gap in most of the studies mentioned above. This technique allows the overall wage gap between men and women to be decomposed into two parts – the first part is due to differences in characteristics across the two sexes, and the second part is due to differences in returns to characteristics across the two sexes (the latter is thought to be a measure of discrimination). The paper by Tansel (2004) estimated the gender wage gap in Turkey across sectors of employment and controlled for selectivity bias, concluding that women are possibly discriminated against in the private sector. Christopoulou and Monastiriotis (2014) examined the wage gap between public and private sectors, concluding that accounting for selection changed the decomposition results (in most cases, the contribution of the endowment effect towards the overall wage gap was lowered). Tiagi (2010) analysed the selectivity-corrected public-private wage gap for men and women separately, stating that the wage premium in the public sector in Canada was predominantly due to differences in characteristics between public and private sector workers.

Studies that have examined the wage gap for Sri Lanka have observed that there is wage premium in favour of men, and this was explained by the discrimination towards women; that is, the contribution of unobservable characteristics towards the overall wage gap (Ajwad and Kurukulasuriya, 2002; Gunerwardena et al., 2008). Gunerwardena et al. (2008) found that women in the Sri Lankan labour force are more educated than men on average; and in addition, the increase in earnings from an extra year of education was higher for women relative to men in the formal sector. However, a persistent wage gap in favour of men was observed in the private sector; whereas in the public sector, women earned more than men on average conditional on observable characteristics. Existing studies have analyzed wage employees, but not the self-employed. Additionally, a limited number of studies on Sri Lanka have focused on the selectivity bias and informal sector. We extend the work of Gunatilaka (2008) who dealt with the selectivity bias and included informal sector employees (but not the self-employed) in the analysis for 2006; however there is no discussion of a significant selectivity issue in the study.

3 Data and descriptive statistics

The analysis for this chapter uses data from the Quarterly Labour Force Survey (QLFS) 2013 conducted by the Department of Census and Statistics, Sri Lanka. The QLFS 2013 is the first in its series to include data from all districts since the end of the war. It is also the first survey to include important variables such as the hours worked and earnings for everyone in the labour force, including the self-employed. The lack of information on the hours worked and not being able to disaggregate private sector workers into formal and informal employment makes the Household Income and Expenditure Survey (2002, 2009/10) a limited data set for this purpose. Previous studies such as the paper by Gunatilaka (2008) for example, were not able to include self-employed individuals. This study aims to give a more detailed analysis of the entire labour force in the country.

Before exploring the summary statistics, it is important to clearly differentiate between formal and informal private sector employment. The definition of informal employment may vary from study to study, often due to data constraints. Several

papers broadly use the definition adopted at the 15th International Conference of Labour Statisticians (ICLS) held in 1993. The criteria used to categorize informal units are as follows: (a) private unregistered enterprises; (b) small number of employees; (c) engaged in non-agricultural activities; (d) non-registration of employees; (e) no formal accounts held. The term "enterprise" comprises of production units employing hired labour, as well as those that are owned and run by individuals who work as self-employed persons with the help of unpaid family members or by themselves. The informal sector therefore includes the number of informal jobs, whether they are carried out in formal sector enterprises, informal sector enterprises, or households. In this study, the informal sector is identified as:

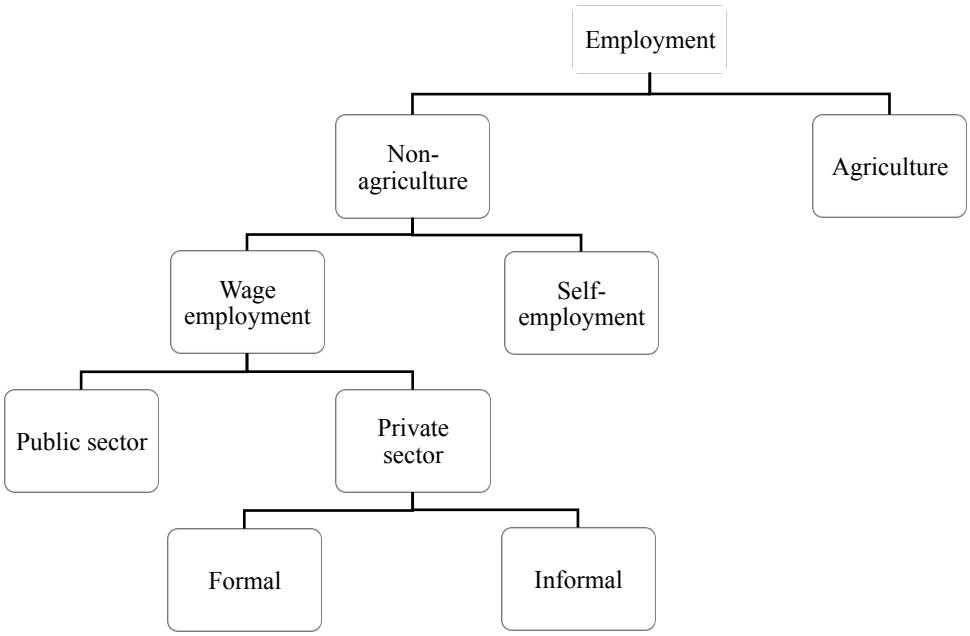
(1) the production units who are registered under the Employees' Provident Fund Act or the Inland Revenue Department who *do not* contribute towards a pension scheme or provident fund on the employees' behalf

(2) the production units who are *not* registered under the Employees' Provident Fund Act or the Inland Revenue Department who either contribute or do not contribute towards a pension scheme or provident fund on the employees' behalf

The Department of Census and Statistics (2013) reported the following characteristics on the formal and informal private sectors in Sri Lanka. 63 per cent of men are in the informal sector whereas 37 per cent are in the formal sector. Over 55 per cent of women are employed in the informal sector and 44 per cent in the formal sector. The level of education has a negative relationship with informal sector participation and a positive relationship with formal sector participation – 23 per cent of individuals working in the informal sector have achieved upper secondary education whereas 80 per cent have achieved only primary education; 77 per cent of individuals working in the formal sector have achieved upper secondary education while 20 per cent have achieved only primary education. The informal sector comprises of primarily own account workers (50 per cent) and employees (34 per cent), while contributing family workers (13 per cent) and employers (3 per cent) are not large numbers in this sector. The formal sector, on the other hand, comprises of mainly employees (90 per cent),

whereas employers (3 per cent), own account workers (5 per cent) and family workers (2 per cent) are negligible.

The above characteristics of the informal sector strengthen the reasoning for separating the informal and formal sectors. As the characteristics of and returns to wage employment may vary from self-employment, this study distinguishes between the two categories. The divisions of employment used in this study can be represented in the following chart:



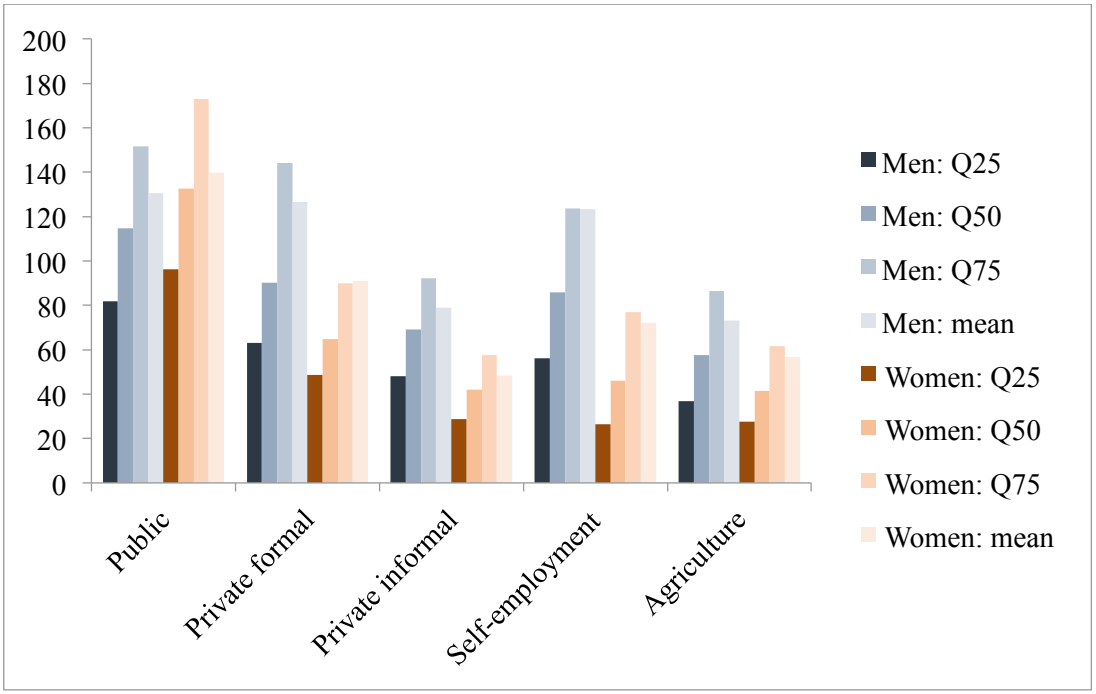
The above figure represents the entire labour force. The *agricultural* workers (both, employees and self-employed individuals) are separated from those working in *non-agricultural* activities such as manufacturing, construction, health and education. The non-agriculture workers are split between the wage employees and self-employed. Within the category of wage employment, the two main sectors are the public and private sectors (the private sector further disaggregated into the formal and informal sectors). Within the category of self-employment, we include all non-agricultural individuals in the formal and informal private sectors who are self-employed.

Table 1 presents descriptive statistics (means) for all individuals in the labour force and the unemployed/not employed, by gender[50]. The earnings per hour are calculated

[50] The descriptive statistics for the full sample are presented in the Appendix Table A1.

as reported weekly earnings divided by the reported actual hours worked during the week[51]. On average, earnings per hour are higher for men relative to women in all sectors, with the exception of the public sector where women earn more than men. For both sexes, earnings are highest in the public and formal private sectors and lowest in the informal private sector and agriculture. By looking at the earnings at various quantiles (the 25th, 50th and 75th) of the distribution, there appears to be a larger gap in earnings across the distribution for men working in the public, formal private sector and the self-employed. For women, the earnings gap across the distribution appears to be the largest in the public sector but not as substantial in the formal private sector and self-employment. The difference in earnings between the 25th and 75th quantiles is relatively small in the informal private sector and agriculture – this is true for both, men and women. Figure 1 is a graphical representation of the earnings at the mean and different quantiles of the distribution for men and women.

Figure1: Earnings per hour at the mean, 25th, 50th and 75th quantiles of the earnings distribution



The earnings are estimated in Sri Lankan rupees (LKR); 1 USD ≈ 146 LKR

[51] Although the average hours worked per week may seem higher in comparison to certain countries, higher working hours have been reported in some developing countries, including Sri Lanka. The LFS 2013 report (DCS) confirms our findings of the hours worked per week.

The Theil index is a generalized entropy inequality measure. Table 1 shows that, in the case of men, 7 per cent of the total inequality is due to between-group inequality (that is, differences in hourly earnings across the public, formal private, informal private, self-employment and agriculture sectors) while the remaining 93 per cent is attributable to within-group inequality (that is, differences in hourly earnings that occurs within each of the five sectors). For women, total inequality is similar to that of men's; however between-group inequality contributes 20 per cent of total inequality whereas within-group inequality contributes 80 per cent. The sub-group inequality is lowest in the public sector, especially for women (0.14) compared to inequality in men's earnings (0.19). The inequality is also low in the informal private sector (0.18 for men and 0.22 for women). Inequality is fairly similar in the formal private and agriculture sectors (approximately 0.40). Some papers claim that earnings in agriculture are more unequal in their distribution because of the inequality of land distribution among the self-employed working in agricultural activities (Gylfason and Zoega, 2002; Frankema, 2006); suggesting that initial land constraints deter growth in agricultural activities thus having an adverse impact on income. The highest level of inequality is seen in the self-employment sector (0.60) – this is perhaps because some entrepreneurial individuals are in this category by choice; others have no alternative and are scrapping a living.

Public and formal private sector employees are more educated compared to individuals in other sectors of employment, on average – this holds for both sexes. Women in the public sector (13.6 years) are more educated than their male counterparts (11.9 years), on average. However, men report higher levels of education in the formal private sector (11.2 years), self-employment (9.4 years) and agriculture (7.1 years) compared to their female counterparts (10.7, 9.3 and 6.5 years respectively). Women employed in the public sector are predominantly in the services industry (for example, health and education). The proportion of women in the manufacturing industry of the formal private sector is high (65 per cent) compared to the proportion of men (46 per cent).

The next set of characteristics observe the skill level of individuals across sectors. Women in the public sector are predominantly high-skilled white collar workers (70 per cent); whereas 40 per cent of men are employed in high-skilled white collar work

in the public sector and 23 per cent in low-skilled blue collar work. The formal private sector employs a larger proportion of white-collar workers compared to the informal sector – this is especially true in the case of men where stark differences are seen across the two sectors. Self-employment contains a somewhat equal split between white collar and blue-collar occupations – for men and women. Agricultural activities are predominantly blue-collar occupations.

Next, the individual characteristics of those who are not in the labour force (that is, unemployed/not employed) will be discussed. These individuals are mainly unmarried (20 per cent of men and 47 per cent of women) and are relatively less educated (approximately 7 years of schooling, on average). Additionally, women who are not employed/unemployed are older, more likely to be married and are more educated than their male counterparts, on average.

Table 1: Descriptive statistics (average) by gender

Variables	Men						Women					
	Non-agriculture				Agriculture	Not in labour force	Non-agriculture				Agriculture	Not in labour force
	Public	Formal private	Informal private	Self			Public	Formal private	Informal private	Self		
Mean	130.66	126.55	78.94	123.29	73.28		139.73	91.16	48.22	72.04	56.79	
25th quantile	81.73	62.94	48.08	56.04	36.92		96.15	48.72	28.84	26.37	27.69	
Median	114.62	90.15	69.23	85.71	57.69		132.69	64.90	41.96	46.15	41.54	
75th quantile	151.65	144.23	92.31	123.63	86.54		173.08	89.74	57.69	76.92	61.54	
Theil's T index:												
Sub-group inequality	0.19	0.35	0.18	0.59	0.37		0.14	0.38	0.22	0.60	0.40	
Total inequality	0.41 (within-group inequality: 0.38; between-group inequality: 0.03)						0.40 (within-group inequality: 0.32; between-group inequality: 0.08)					
Individual characteristics:												
Hours worked per week	47.1	50.9	46.8	49.2	39.7		40.0	47.7	43.4	38.0	34.2	
Years of age	41.6	35.7	37.7	44.4	46.7	26.1	39.6	32.8	38.3	44.9	46.3	29.6
Urban	0.21	0.26	0.19	0.23	0.04	0.17	0.20	0.20	0.20	0.18	0.01	0.17
Sinhalese	0.81	0.84	0.69	0.73	0.70	0.69	0.82	0.87	0.82	0.80	0.68	0.69
Married	0.86	0.69	0.69	0.86	0.86	0.20	0.76	0.53	0.48	0.72	0.71	0.47
Years in education	11.88	11.24	8.97	9.40	7.10	6.70	13.58	10.73	9.05	9.30	6.46	7.72
Industry (base: services)												
Manufacturing	0.07	0.46	0.44	0.27			0.03	0.65	0.49	0.47		
Occupation (base: other)												
High skilled white collar	0.40	0.35	0.07	0.21	0.01		0.70	0.25	0.12	0.23	0.01	
Low skilled white collar	0.25	0.22	0.20	0.22			0.22	0.16	0.20	0.23	0.01	
High skilled blue collar	0.05	0.17	0.30	0.30	0.63		0.01	0.27	0.28	0.45	0.39	
Low skilled blue collar	0.23	0.26	0.43	0.27	0.36		0.06	0.32	0.40	0.09	0.59	
Number of observations	2,255	1,836	3,394	4,379	5,601	14,810	1,798	1,203	1,261	1,676	1,979	29,061

The earnings are estimated in Sri Lankan rupees (LKR); 1 USD ≈ 146 LKR

4 Methods

This section looks at the techniques used to address some of the econometric issues relating to the estimation of earnings functions of individuals in different sectors of employment. The sectors are categorized as follows – non-agriculture workers in the public sector, private (formal and informal) wage employment and self-employment, and agriculture workers.

The baseline model used to estimate the earnings function for individual i in sector j utilizes the Mincerian approach:

(1) $$ln\,(Y_{ij}) = \beta_0 + \beta_1 S_{ij} + \beta_2 Z_{ij} + v_{ij}$$

where the dependent variable $ln(Y_{ij})$ is the natural log of earnings per hour of individual i in sector j, S_{ij} is his/her level of schooling, Z_{ij} is a vector of other observed determinants of earnings (including age, gender, ethnic background, education) and v_{ij} is an error term representing the unobserved determinants.

As discussed in the literature, two main concerns arise with regard to estimating the earnings functions – sample selection and possible endogeneity of the education variable – which will be the focus of this chapter. Ignoring these concerns will lead to biased estimates.

4.1 Sample selection

To estimate a "wage offer" equation for people of working age, the ideal would be to include all individuals whether or not they are working at the time of the survey. However, wages or earnings are only observed for people who are in the labour force at the time of the survey, leading us to use a selected sample. People self-select into employment, hence whether the wage is observed or not will depend on an individual's labour supply decision. Failing to account for this (that is, sample selection) may lead to biased estimates of equation (1). Additionally, given that women are less likely to be employed than men in Sri Lanka, the selectivity bias

could affect the comparisons of returns to men's and women's education – for example, the small proportion of women in the labour force may be relatively more able or ambitious and therefore their unobservable characteristics could be positively correlated with schooling and wages. To deal with the issue of potential sample selection bias, this chapter employs Lee's (1983) two-stage method.

In the first stage, the probability of labour force participation is estimated using a multinomial logit (MNL). The decision to participate in the labour market or not depends on human capital characteristics (for example, education, age, marital status which influence the offer wage) and household characteristics (for example, the presence of children/elders that may influence the reservation wage through its effect on household productivity and the demand for leisure). An individual will choose to enter the labour market if the offered wage is greater than the reservation wage. We are faced with the necessity to distinguish individuals in the work force by sector of employment. The following sectors were identified - non-agriculture (public wage, private formal wage, private informal wage, and self-employment) and agriculture sectors. By assuming that employment outcomes are the result of stochastic utility maximization, the probability of employment in any particular sector can be justified. The utility of being in a particular sector j (U_{ij}) can be thought of as a linear function of education (S_{ij}), other observable characteristics (Z_{ij}) and error terms (η_{ij}):

(2) $$U_{ij} = \gamma_0 + \gamma_1 S_{ij} + \boldsymbol{\gamma_2 X_{ij}} + \eta_{ij}$$

In the second stage of Lee's two-stage estimation, the probabilities of individual i being in sector j are taken from the first stage MNL estimation to construct inverse Mills ratios (λ_{ij}) which are then included in the earnings functions to correct for sample selection:

(3) $$\lambda_{ij} = \frac{\varphi(h_{ij})}{\Phi(h_{ij})} \quad \text{and} \quad h_{ij} = \Phi^{-1}(P_{ij})$$

where $\varphi(h_{ij})$ and $\phi(h_{ij})$ are the respective density and cumulative distribution functions of the standard normal distribution.

Equation (1) is modified to account for selectivity and thus give consistent estimates of the earnings functions[52]:

(4) $$ln\,(Y_{ij}) = \beta_0 + \beta_1 S_{ij} + \boldsymbol{\beta_2 Z_{ij}} + \tau_{ij}\lambda_{ij} + \zeta_{ij}$$

where ζ_{ij} is the error term.

If the unobservables determining selection into employment are independent of the unobservables determining the earnings (per hour), we can conclude that there is no sample selection bias.

Earnings are one factor determining the utility from employment; hence the determinants of earnings ($\boldsymbol{Z_{ij}}$) will be included along with the determinants of employment ($\boldsymbol{X_{ij}}$) in the MNL. In order to identify the participation model, variables that determine employment but not the earnings should be included. Identification can be achieved by the exclusion of several individual and household characteristics from the earnings functions (spouse/parent's employment status, proportion of children below the age of 6/between the ages of 6 and 18/proportion of adults over the age of 65 in the household)[53].

The literature suggests that women's participation decisions are perhaps more strongly determined by factors such as marital status and the presence of young children, compared to the impact that these factors have on a man's decision to participate or not. To deal with potential sample selection bias, it is important to have at least one exclusion restriction – a variable that affects choice of employment, but not the earnings. As the Wald tests suggest appropriateness of the exclusion restrictions, even if a single variable, for example the proportion of children, might not be a good criterion when controlling for selection in the case of men, we chose to

[52] As probabilities obtained from the MNL are inserted into the earnings function manually, we correct the standard errors by bootstrapping in the earnings function in order to get asymptotically consistent values.

[53] The appropriateness of this identification was tested using Wald tests of joint significance of the identifying variables in the MNL and earnings functions in each sector of employment. The tests suggested that the exclusion restrictions were appropriate.

proceed as the Wald tests suggest that there is at least one appropriate exclusion restriction.

The coefficients on the gender dummy variable (female) in the participation model for the aggregate sample are large in magnitude and highly significant[54]. In order to allow for gender-differences in employment choices and earnings, all estimations are carried out for men and women separately.

An important property of the MNL model is the "Independence of Irrelevant Alternatives" (IIA) – the relative odds of being in one sector should be independent of the relative odds of being in alternative sectors. Under IIA, the restricted estimator where one alternative is ignored would still be consistent, yet inefficient. Hausman tests conducted for each sector suggested that the IIA assumption was not violated in most cases, with the exception of the informal sector and self-employment - a possible explanation for this could be some extent of similarity in the characteristics of individuals who join these two sectors. However, we are interested in identifying the effect of selection into a particular sector on earnings. Bourguigon et al. (2007) state that the "selection bias correction based on the multinomial logit model provides a fairly good correction for the outcome equation even when the IIA hypothesis is violated" (pp. 199 – 200) based on Monte Carlo simulations. Alternative approaches would allow the IIA assumption to be relaxed (for example, the conditional logit or nested logit); however, these approaches require alternative-specific variables for all alternatives and not just for the chosen alternative that are not available in this data set. The significance or insignificance of the selectivity correction terms in the earnings functions will suggest whether the hypothesis of correlation between unobservable characteristics determining employment and earnings can be rejected or not – that is, whether sample selectivity affects earnings.

4.2 Endogenous explanatory variable(s)

In addition to the potential bias of sample selectivity, this chapter also deals with the possible endogeneity of education - that is, an individual's level of education could be

[54] The results are presented in the Appendix Table A3

correlated with the error term (ζ_{ij}) in equation (4) because of unobservable individual heterogeneity. From chapter 2, it was concluded that it is important to take into account the possible endogeneity of education. Failing to do so could lead to biased Ordinary Least Squares (OLS) estimates. To address this issue, an Instrumental Variable (IV) technique was employed, which involved the use of variables that are correlated with education but uncorrelated with earnings. Such variables include parents' and spouse's education levels and a change in the compulsory years of schooling. In this chapter to deal with the issue of endogeneity whilst controlling for selectivity, a control-function approach is adopted (see Garen, 1984; Wooldridge, 2005; Söderbom et al., 2006, Kuépié et al., 2009).

The method modifies Lee's (1978) sample selection model. Schooling (S_{ij}) is measured using instruments (I_{ij}) and other observable characteristics (Z_{ij}):

$$\hat{S}_{ij} = \delta_0 + \delta_1 I_{ij} + \delta_2 Z_{ij} + \varepsilon_{ij} \tag{5}$$

where $\hat{S}_{ij}$ is the predicted level of schooling of individual i in sector j, I_{ij} is a vector of instruments, Z_{ij} is a vector of observable characteristics such as age, ethnicity, gender and marital status, and ε_{ij} is the error term. Endogeneity of education arises if ε_{ij} is correlated with ζ_{ij} (the error term from equation 4).

The residuals from equation 5 are obtained and inserted into the MNL (equation 2) and the sectoral earnings equations (equation 4). Therefore, the MNL is modified as follows:

$$U_{ij} = \gamma_0 + \gamma_1 S_{ij} + \gamma_2 Z_{ij} + \gamma_3 \vartheta_{ij} + \omega_{ij} \tag{6}$$

where ϑ_{ij} represent the residuals obtained from the reduced-form education equation (equation 5) and ω_{ij} is the error term[55].

[55] Standard errors have been bootstrapped in both, the MNL and earnings functions.

The earnings functions are modified to include the residuals from the reduced-form education equation and the inverse Mills ratios from the MNL:

$$ln\,(Y_{ij}) = \beta_0 + \beta_1 S_{ij} + \beta_2 Z_{ij} + \tau_{ij}\lambda_{ij} + \kappa_{ij}\vartheta_{ij} + \mu_{ij} \quad (7)$$

where μ_{ij} is the error term.

This approach produces consistent estimates of the parameters of interest provided the standard conditions for identification hold, and provided the instruments (I_{ij}) are independent of the residuals obtained from equation 5 (ϑ_{ij}) and uncorrelated with the residual of the earnings function (equation 7). The two-stage least squares approach, on the other hand, does not require independence between the residuals and the unobservable component of the earnings function, but requires zero covariance (Söderbom et al., 2006; Wooldridge, 2005). The control function approach can address the sample selectivity problem provided the instruments (I_{ij}) are independent of the error term for the selected sample.

Equation 7 accounts for the selectivity into a sector (with the inclusion of selectivity correction terms from the MNL – λ_{ij}) and the endogeneity of education (residuals from the reduced form - ϑ_{ij}). It is thus the key equation of interest. Z_{ij} is a vector of individual characteristics. These variables are also included in the probability of employment model and the reduced-form education equation. *Age*, *age-squared* and *age-cubed* are variables included in the regression in order to capture the non-linear relationship that age is believed to have with employment choice and earnings. Dummy variables are used to control for the individual being from an urban or rural area *(=1 if individual is from an urban area, 0 otherwise)*, and ethnic background (*=1 if Sinhalese, or 0 otherwise)*, marital status *(=1 if married, 0 otherwise)*.

The next set of variables affect the probability of selection into a particular sector of employment (or not working) and/or the individual's level of education, but will have no effect on earnings[56]. These include the spouse/mother/father's level of education

[56] As mentioned earlier, the appropriateness of this identification was tested using Wald tests of joint significance of the identifying variables in the MNL/reduced form of education, and insignificance in the earnings functions (for each sector of employment).

and sector of employment (public, private or not in the labour force), proportion of children in the household below the age of 6, proportion of children between the ages of 6 and 18 (inclusive) and the proportion of elders over the age of 65. Wooldridge (2010) suggests using the same set of instruments to deal with both issues (selectivity and endogeneity) for the following reasons. Having at least one instrument that "primarily" affects selection but not education, and one other instrument that affects education but not selection forces discipline on the procedure when we are faced with two forms of potential bias. However, Wooldridge (2010) states that the assumptions made in this model allow for the possibility to have the same set of instruments to deal with both forms of bias. Having the same set of instruments means that the reduced form for education tends to suffer from collinearity since the IMR will be a function of the same variables. However as Wooldridge (2010) discusses, we are not interested in the reduced form parameters and therefore the collinearity introduced by having the same instruments appearing linearly as those appearing in the IMR is not of much concern. More importantly, by choosing instruments that could potentially affect education but not selectivity and vice versa, we are making exclusion restrictions in the selection equation – violation of these restrictions can lead to inconsistent wage estimates. Therefore, best practise is to have a vector of all exogenous variables in the reduced-form education equation and MNL. In the next section, the main results controlling for sample selectivity and endogeneity of education will be presented.

5 Results

The results are presented as follows. Section 5.1 analyses the determinants of sector of employment by gender. Section 5.2 explores the earnings functions for men and women working in each sector of employment, after controlling for endogeneity of education and sample selectivity. The reduced-form education estimates (by gender and for the full sample) are presented in Appendix 1, along with the probability of employment model and earnings functions for the full sample. Section 5.3 analyses the predicted earnings in different employment sectors, by level of education and age.

5.1 Determinants of sector of employment

Tables 3 and 4 present the estimates obtained from the multinomial logit (MNL) on the probabilities of employment for men and women respectively[57]. Men working in the public and formal private sectors report the highest levels of education, compared to other sectors of employment – the coefficients are 0.173 and 0.170 respectively. The pattern for women is markedly different. The coefficient on education for women in the public sector is much greater (0.713), and is insignificant for women in the formal private sector. This suggests that more educated women end up in the public sector, whereas more educated men end up in both, the public and formal private sectors. For both genders, the less educated are more likely to be found in the informal and agriculture sectors. This result confirms the fact that the informal sector and agriculture employment attracts less-educated individuals. The signs of the education coefficients are different for self-employed men and women – less educated men are likely to be in self-employment in contrast to more educated women.

The probability of labour force participation in all sectors of employment increases with age; the negative sign on the age-squared coefficient indicates that the acquisition of human capital slows down after a certain age. Men from urban areas are not surprisingly more likely to join the formal private sector or self-employment, whereas men in rural areas are more likely to join agricultural employment. Women from urban areas are likely to join the formal and informal private sectors, and women from rural areas are likely to join the public sector and agricultural activities. Individuals from a Sinhalese (ethnic) background are more likely to be employed in all sectors compared to their non-Sinhalese counterparts, especially in the formal private sector. Married men are more likely to participate in the labour force, whereas married women are less likely to participate.

[57] A Likelihood Ratio test of the null of equality in coefficients between any two sectors is rejected at the 1 per cent significance level for the full sample and the sub-samples by gender. This suggests the suitability of distinguishing the sectors rather than considering them at an aggregate level, that is, public versus private (formal and informal wage and self-employed). In order to deal with the potential selectivity bias, multiple exclusion restrictions were employed in the MNL. A likelihood ratio test was used to test the relevance of these restrictions. The null hypothesis of coefficients simultaneously being equal to zero was rejected at the 1 per cent significance level (for all samples).

For men, the policy variable capturing the effect of a change in years of compulsory schooling in 1997 has a positive coefficient; this indicates that the increase in compulsory schooling has had a positive impact on men's probability of employment. The change in compulsory schooling laws has had no effect on a woman's probability of employment in the public sector – a possible explanation for this is that women in this sector have attained high levels of education and a change in schooling laws had no effect on their participation choices. The change in compulsory schooling laws has had a positive effect on female employment in the formal and informal privates sectors. Spouse's education is an important determinant of labour force participation for men – the more educated the spouse is, the greater the probability of employment. However for women, spouse's education is not a determinant of employment with the exception of the private formal sector and agriculture – in the formal sector, spouse's education increases the probability of employment in that sector whereas in agricultural employment the spouse's level of education acts as a deterring factor against working in that sector. Parents' education is included in the participation model because it is expected that more educated parents will have access to networks and assist with job search if they have more social capital. Parent's education decreases the probability of employment in certain sectors, except for the formal private sector for both genders and the public sector for women where it has no significant effects.

The spouse's sector of employment has different effects on the probabilities of labour force participation of both sexes. Having a spouse working in the public sector increases the probability of an individual participating in the public sector himself/herself while having a spouse working in the private sector decreases own probability of public sector employment – this is true for both sexes. Having a spouse who is not in the labour force would increase the probability of a woman entering the informal private sector, and increase the probabilities of men entering the informal private, self-employment and agriculture sectors.

The presence of children below the age of 6, or between the ages of 6 and 18 in a household has no effect on men's choice of employment sector (apart from a deterring effect on the probability of formal private sector employment). However, the impact on women's labour force participation decisions is different; the presence of children

below the age of 6 or between the ages of 6 and 18 in the household reduces the probability of participation in the public, formal and informal private sectors, and the presence of children of school age (between ages 6 and 18) reduces the probability of employment in self-employment and agriculture. This suggests that women are more likely to stay at home and look after children. The presence of elders (above the age of 65) in the household increases the probability of employment in the informal sector, self-employment and agriculture for both sexes; and additionally it increases women's probability of employment in the formal private sector.

The analysis is not able to account for the non-pecuniary elements for enumeration such as job security, work effort and compensating differentials which could encourage individuals to move from one sector to another. Several studies (for example, Mengistae, 1999) state that individuals prefer the public sector because of job security and fixed job contracts. This was not the case in other sectors such as agriculture and the informal sector – thus, wages do not explain everything. Individuals with low skill levels/education/experience may prefer the public sector, but individuals with more skills/experience/education would prefer sectors where wages are determined by productivity. However by controlling for selectivity, this analysis implicitly tries to address this issue.

Having looked at determinants of the sector of employment by gender, the next section will analyse the earnings functions for each sector after controlling for sample selectivity and the endogeneity of education.

Table 3: Determinants of probability of employment (base category: unemployed/not employed) for men

Variables	Non-agriculture								Agriculture	
	Public		Private wage (formal)		Private wage (informal)		Self-employed			
	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic
Individual characteristics:										
Education (years)	0.173***	4.09	0.170***	3.78	-0.161***	-5.32	-0.069**	-2.37	-0.308***	-12.13
Residuals from first stage	0.181***	4.05	0.054	1.14	0.149***	4.74	0.101***	3.33	0.201***	7.74
Age	1.148***	10.64	1.324***	15.66	1.210***	25.26	1.397***	25.61	1.192***	26.04
Age^2	-0.018***	-7.35	-0.025***	-13.71	-0.024***	-24.31	-0.026***	-24.73	-0.022***	-23.89
Age^3	$5.97x10^{-5}$***	3.40	$1.35x10^{-4}$***	10.30	$1.31x10^{-4}$***	21.22	$1.43x10^{-4}$***	22.46	$1.12x10^{-4}$***	19.89
Urban	-0.021	-0.25	0.315***	3.76	0.132*	1.92	0.195***	3.01	-1.564***	-17.92
Sinhalese	0.412***	5.35	0.732***	9.01	0.114**	2.02	0.076	1.35	0.171***	3.23
Married	0.555***	3.12	0.683***	4.17	0.570***	4.43	0.479***	3.67	0.570***	4.81
Policy (=1 if age ≤ 26)	1.329***	5.44	1.316***	5.93	1.499***	9.43	1.253***	7.23	1.737***	10.90
Education:										
Spouse	0.106***	5.59	0.067***	3.62	0.070***	5.02	0.115***	8.39	0.087***	7.04
Mother	-0.043**	-2.52	-0.005	-0.31	-0.059***	-4.71	-0.031**	-2.23	-0.092***	-6.67
Father	-0.065***	-3.39	-0.014	-0.80	-0.069***	-4.79	-0.045***	-2.77	-0.156***	-8.95
Sector of employment:										
Spouse: in public sector	0.421***	3.21	-0.334**	-2.16	-1.271***	-7.38	-0.855***	-6.66	-0.408***	-3.23
in private sector	-0.451***	-4.39	0.082	0.79	-0.219**	-2.53	-0.696***	-8.31	-0.771***	-9.75
Mother: in public sector	-0.399	-1.33	-0.502**	-2.18	-1.351***	-4.33	-1.036***	-3.11	-0.128	-0.48

in private sector	0.130	0.84	0.084	0.67	0.149	1.53	0.062	0.49	0.106	0.90
Father: in public sector	0.358	1.50	-0.435**	-2.13	-0.630***	-2.87	-0.678***	-2.58	-0.480*	-1.84
in private sector	-0.091	-0.68	-0.326***	-2.92	0.069	0.71	-0.140	-1.28	-0.367***	-3.27
Household characteristics:										
Proportion of children (age<6)	0.490	1.54	-0.689**	-2.05	-0.415	-1.47	0.410	1.55	0.234	0.92
Proportion of children (6≤age≤18)	0.320	1.44	-0.667***	-2.71	-0.305	-1.60	0.299*	1.67	0.112	0.66
Proportion of elders (age>65)	0.214	0.59	0.342	0.91	0.439**	2.01	0.556***	3.38	1.086***	7.86
Constant	-22.580***	-15.75	-23.391***	-22.03	-16.446***	-27.66	-21.324***	-27.70	-15.092***	-24.29
N = 32, 275; LR $chi^2(115)$ = 33643.66; Prob>chi^2 = 0.00; Pseudo R^2 = 0.34										

Controlled for missing values of spouse and parent(s) employment sector and/or level of education; * p<0.1, ** p<0.05, ***p<0.01; Default categories are rural, non-Sinhalese, unmarried, spouse/mother/father being unemployed/not employed.

Table 4: Determinants of probability of employment (base category: unemployed/not employed) for women

Variables	Non-agriculture								Agriculture	
	Public		Private wage (formal)		Private wage (informal)		Self-employed			
	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic
Individual characteristics:										
Education (years)	0.713***	17.92	-0.023	-0.57	-0.184***	-5.66	0.056**	2.09	-0.253***	-11.76
Residuals from first stage	-0.069*	-1.68	0.088**	2.12	0.163***	4.80	-0.006	-0.23	0.150***	6.63
Age	0.258**	2.26	1.033***	14.61	0.920***	15.94	0.600***	8.96	0.789***	13.26
Age^2	0.001	0.20	-0.021***	-13.17	-0.017***	-13.65	-0.010***	-7.35	-0.013***	-11.13
Age^3	$5.11x10^{-5}$**	-2.55	$1.14x10^{-4}$***	10.42	$8.67x10^{-5}$***	10.57	$4.64x10^{-5}$***	5.37	$6.39x10^{-5}$***	8.16
Urban	-0.415***	-4.88	0.357***	4.24	0.431***	5.32	0.050	0.70	-2.814***	-12.07
Sinhalese	-0.047	-0.52	1.219***	11.69	1.155***	12.90	0.547***	7.30	0.451***	7.06
Married	-0.604***	-5.50	-0.990***	-8.74	-1.027***	-9.65	-0.596***	-6.38	-0.351***	-4.05
Policy (=1 if age ≤ 26)	-0.178	-0.86	1.077***	6.25	1.700***	10.05	0.052	0.26	1.054***	5.46
Education:										
Spouse	0.002	0.11	0.035**	2.15	0.004	0.27	-0.016	-1.31	-0.047***	-4.25
Mother	-0.003	-0.16	0.020	1.19	-0.034**	-2.10	-0.048**	-2.18	-0.163***	-6.45
Father	-0.034	-1.57	-0.012	-0.67	-0.062***	-3.37	-0.017	-0.61	-0.032	-0.99
Sector of employment:										
Spouse: in public sector	0.843***	6.81	-0.377**	-2.13	-0.542***	-2.77	-0.062	-0.49	0.655***	5.87
in private sector	-0.284**	-2.49	-0.157	-1.36	-0.376***	-3.53	-0.041	-0.49	-0.358***	-4.78
Mother: in public sector	-0.096	-0.33	-0.879***	-2.81	-0.253	-0.83	0.383	0.86	1.764***	4.66

in private sector	0.054	0.31	0.102	0.84	0.178	1.39	0.335*	1.68	0.175	0.75
Father: in public sector	0.348	1.30	-0.600**	-2.51	-0.050	-0.21	-0.237	-0.52	-0.799	-1.38
in private sector	0.019	0.13	-0.271**	-2.39	-0.107	-0.86	0.086	0.45	-0.497**	-2.19
Household characteristics:										
Proportion of children (age<6)	-0.519*	-1.88	-2.679***	-7.68	-2.057***	-5.91	-0.186	-0.78	-0.319	-1.40
Proportion of children (6≤age≤18)	-0.304*	-1.65	-1.216***	-5.59	-0.587***	-3.05	0.573***	3.95	0.477***	3.52
Proportion of elders (age>65)	-0.127	-0.36	1.241***	3.35	0.818***	3.27	0.884***	5.10	1.194***	7.64
Constant	-16.585***	-11.15	-17.583***	-20.96	-14.818***	-20.93	-13.276***	-13.41	-12.458***	-13.96

N = 36, 978; LR $chi^2(115)$ = 14952.13; Prob>chi^2 = 0.00; Pseudo R^2 = 0.24

Controlled for missing values of spouse and parent(s) employment sector and/or level of education; * p<0.1, ** p<0.05, ***p<0.01; Default categories are rural, non-Sinhalese, unmarried, spouse/mother/father being unemployed/not employed.

5.2 Earnings across sectors

Having looked at the variables that affect the probability of employment in various sectors, the estimates of the earnings functions will be discussed. Note that a reduced form of education regressed on all exogenous variables (instruments and observables) was estimated[58], and the residuals included in the earnings functions; therefore the earnings functions have controlled for endogeneity of education – ignoring this would lead to downward biased OLS estimates. The selectivity correction terms are also included in the earnings functions, thus controlling for any form of selectivity. The earnings functions estimated by gender are presented in Tables 5 and 6.

The returns to education are higher for women in every sector of employment relative to men, with the exception of agriculture where the returns for men are 5.4 per cent. For men and women, the highest returns are seen in the formal private sector – 16.2 and 16.8 per cent, respectively. The returns to education differ greatly across men and women in the public sector - an extra year of schooling increases earnings by 8.8 per cent for men and by 14.1 per cent for women. In self-employment, the returns to education are 12.6 per cent for women and 9.3 per cent for men. Hamilton (2000) suggested that there are substantial non-pecuniary benefits of self-employment. He discussed that this sector has a spectrum of individuals – some are experienced, skilled and enjoy high returns whereas others have lower skill levels and low returns. Additionally, Parker (2009) stated that the return to self-employment includes a return to capital and a return to labour. These two viewpoints are possible explanations for the high returns to self-employment that we have noted above. The returns are lowest in the informal private sector – 4.3 and 5.8 per cent for men and women, respectively[59].

[58] Refer to reduced form estimates in the Appendix (Tables A2, A5 and A6 for the full sample, men and women respectively)

[59] The earnings functions including controls for occupation type are reported in the Appendix (Tables A7 and A8 for men and women respectively). In all sectors, with the exception of the returns to education for men in agriculture where the coefficient rises slightly from 0.54 to 0.56 after controlling for occupation and the informal sector for men where the returns are similar in both scenarios, there is a fall in the returns to education after including the controls for occupation type. The most substantial fall is seen in the formal private sector, for both, men and women – falling from 0.162 to 0.122 for men, and from 0.168 to 0.114 for women. This brings the returns to education in the private formal sector closer to the returns in the public sector for both sexes. For women after controlling for occupation

The returns are higher for individuals from urban areas in the private formal and informal sectors and self-employment. However there are no significant differences in earnings across rural and urban areas for women in the public sector and agriculture; whereas for men, differences across urban and rural areas in these two sectors of employment are significant only at the 10 per cent level. The earnings functions suggest that Sinhalese men receive lower returns compared to their non-Sinhalese counterparts in all sectors apart from the public sector. However, the participation model implied that Sinhalese are more likely to join the labour force; thus, the Sinhalese who join the labour force receive lower earnings in all sectors with the exception of the public sector. Married men receive higher returns in all sectors with the exception of the public sector, compared to their unmarried counterparts. No such differences are seen across married and unmarried women.

Since education is likely to suffer from omitted variable bias in the form of unobserved characteristics, a control function approach is used to check if education is in fact, endogenous. First, a reduced form regression is run with education as the dependent variable and all exogenous variables (including the instruments and other explanatory variables). An F-test to check for the joint significance of the coefficients on all instruments used in the reduced form for education is conducted. For all sub-samples, the hypothesis that these coefficients are jointly equal to zero was rejected. The estimates of the reduced form, along with the F-tests, are presented in Appendix 1 (Tables A2, A5 and A6). Next, the residuals are obtained from the reduced form equation. These residuals are included in the MNL and structural equations (wage functions). By looking at the significance of the residuals in the structural equations, we can examine whether education should be treated as exogenous or endogenous. The null hypothesis is that coefficient on the residuals is zero and that education is therefore exogenous. This hypothesis was rejected at the 1 per cent significance level for the private formal and self-employment sectors for both sexes and for women in

type, the returns in the formal sector (0.114) are now lower than the returns in the public sector (0.126) whereas exclusion of the occupation dummies suggested that returns are higher in the formal sector (0.168, as opposed to 0.141 in the public sector). However, these controls were not included in the reduced form and MNL because an individual may choose the employment sector and occupation simultaneously and so it would be imprecise to predict, say, the probability of employment in the public sector given that he/she is a white collar worker; therefore it was decided to leave them out of the main analysis.

the public sector. The null was rejected at 5 per cent significance level for male agriculture workers and rejected at the 10 per cent level for male public sector workers. Therefore, there is evidence to suggest that education is endogenous in these sectors. In the cases of men working in the private informal sector, and women working in the informal or agriculture sectors, the OLS estimates of the earnings functions would suffice where education is treated as exogenous.

The selectivity correction term is insignificant at the 1 per cent significance level in most cases, for both men and women. This suggests that the hypothesis of correlation between unobservables affecting employment and earnings can be rejected – that is, sample selection does not lead to biased estimates. However for men in the formal private sector, selectivity has a positive impact on earnings – unobservable factors associated with preference towards this sector are correlated with the residual of the earnings function. The sectoral wage equations for the full sample are presented in Appendix 1 (Table A4). A dummy variable was used in the estimation to account for gender differences. This variable yielded a negative coefficient suggesting that women are disadvantaged in terms of returns – excluding the public sector, men receive higher returns than women.

As the main results from Tables 5 and 6 suggest that the coefficients on the earnings functions differ across men and women, a Blinder-Oaxaca decomposition would enable us to explore these differences further. A decomposition of the gender wage-gap across sectors is useful to identify the contribution of differences in male-female characteristics and differences in returns to characteristics towards the overall wage gap. This chapter employed a three-stage model and identified two forms of bias, where the third stage (estimation of the earnings functions) depend on the first and second stages (obtaining the residuals for education and the multinomial logit, respectively). However, the decomposition uses the estimates for the first and second stages of the full sample after which, the third stage is estimated for the sub-samples by gender. The decomposition technique and results are presented in Appendix 2. The key findings are as follows. Women earn more than men on average in the public sector, and this is predominantly due to differences in characteristics (such as education) between the two genders. In the other sectors, the unexplained component

(which is a measure of unobservable characteristics and discrimination towards women) has a greater impact on the wage gap in favour of men.

Table 5: Earning functions for men

Dependent variable: log(earnings per hour)	Non-agriculture								Agriculture	
	Public		**Private wage (formal)**		**Private wage (informal)**		**Self-employed**			
	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic
Education (years)	0.088***	4.69	0.162***	11.94	0.043***	3.48	0.093***	8.02	0.054***	3.97
Age	0.117	1.57	0.070*	1.70	0.119***	4.54	0.097**	1.97	0.054*	1.67
Age^2	-0.002	-1.17	-0.001	-0.92	-0.002***	-4.11	-0.001	-1.38	-0.001	-1.62
Age^3	$1.21x10^{-5}$	0.90	$1.42x10^{-6}$	0.20	$1.44x10^{-5}$***	3.53	$5.00x10^{-6}$	0.71	$7.44x10^{-6}$	1.43
Urban	0.091*	1.90	0.225***	4.72	0.172***	5.43	0.252***	4.68	0.226*	1.70
Sinhalese	-0.026	-0.68	-0.082*	-1.77	-0.068**	-2.04	-0.143***	-3.60	-0.232***	-7.23
Married	-0.055	-1.06	0.197***	3.00	0.155***	3.08	0.144*	1.91	0.184***	3.15
Residuals from first stage	-0.023*	-1.86	-0.043***	-3.29	-0.016	-1.25	-0.032***	-2.84	-0.021**	-2.13
Lambda (selectivity correction)	-0.020	-0.26	0.199**	2.57	0.015	0.14	0.157	1.31	-0.047	-0.38
Constant	1.589	1.57	0.993	1.47	1.866***	4.24	1.390	1.63	2.847***	6.03
Number of observations	2,255		1,836		3,394		4,379		5,601	
R^2	0.11		0.18		0.04		0.06		0.03	

* p<0.1, ** p<0.05, ***p<0.01; Default categories are rural, non-Sinhalese, unmarried; the standard errors are bootstrapped with 500 replications

Table 6: Earnings functions for women

Dependent variable: log(earnings per hour)	Non-agriculture								Agriculture	
	Public		**Private wage (formal)**		**Private wage (informal)**		**Self-employed**			
	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic
Education (years)	0.141***	5.40	0.168***	9.02	0.058**	2.51	0.126***	6.23	0.014	0.49
Age	0.107	0.87	0.051	0.62	0.011	0.26	0.120	1.34	0.045	0.47
Age^2	-0.001	-0.50	-0.001	-0.34	$-1.16x10^{-4}$	-0.12	-0.002	-1.04	-0.001	-0.47
Age^3	$7.57x10^{-6}$	0.32	$2.71x10^{-6}$	0.21	$5.19x10^{-8}$	0.01	$8.38x10^{-6}$	0.78	$5.45x10^{-6}$	0.47
Urban	-0.004	-0.06	0.206***	2.99	0.307***	5.94	0.303***	3.15	-0.129	-0.24
Sinhalese	0.028	0.48	-0.255*	-1.90	-0.022	-0.25	-0.152	-1.55	0.044	0.64
Married	-0.011	-0.20	-0.039	-0.30	-0.070	-0.99	-0.012	-0.12	0.072	0.95
Residuals from first stage	-0.052***	-4.26	-0.097***	-5.50	-0.027	-1.10	-0.065***	-3.04	-0.008	-0.36
Lambda (selectivity correction)	0.051	0.55	0.145	0.57	0.064	0.47	0.382	1.58	-0.024	-0.11
Constant	0.599	0.33	1.269	0.82	2.745***	4.12	0.466	-0.26	2.747*	1.70
Number of observations	1,798		1,203		1,261		1,676		1,979	
R^2	0.18		0.11		0.04		0.05		0.01	

* p<0.1, ** p<0.05, ***p<0.01; Default categories are rural, non-Sinhalese, unmarried; the standard errors are bootstrapped with 500 replications

5.3 Predicted earnings

Having analysed the earnings functions, the average earnings per hour an individual will receive in a specific sector conditional on observable characteristics are examined. Table 7 presents the predicted earnings for men and women in different sectors[60]. For women, the highest expected wages per hour are in the public sector whereas for men it is in self-employment. Women earn more than men on average in the public sector. The formal private sector pays fairly similar wages compared to the public sector for men (149 LKR in the public sector and 141 LKR in the private sector), but there is a large gap in earnings for women (159 LKR in the public sector and 117 LKR in the formal private sector). The informal sector pays the lowest earnings for both genders; this is especially true for women (57 LKR for women and 104 LKR for men).

Table 7: Predicted (average) earnings per hour by gender

	Men	**Women**
Public	148.97 (2255)	159.40 (1798)
Private formal	141.03 (1836)	116.56 (1203)
Private informal	104.46 (3394)	57.41 (1261)
Self-employment	168.23 (4397)	102.92 (1676)
Agriculture	105.13 (5601)	77.66 (1979)

The number of observations is in parentheses; the earnings are estimated in Sri Lankan rupees (LKR); 1 USD ≈ 146 LKR

In order to understand the differences in earnings for younger (age≤30) and older (age>30) individuals, the sample is split by age. Table 8 presents the predicted earnings. As one would expect, age has a positive relationship with predicted earnings – with older individuals earning more, on average, than younger individuals in all sectors with the exception of younger men in agriculture and younger women in the informal sector earning more than their older counterparts. However, the two exceptions display minor differences. The largest differences across old and young workers are seen for women in the public sector - older women receive considerably higher wages (174 LKR) compared to younger women (113 LKR). The average wage gap by age is similar between the public and formal private sectors for men.

[60] The predicted earnings do not differ substantially if we control for sample selectivity or not. Therefore, predicted earnings without the sample correction were used.

Table 8: Predicted (average) earnings per hour - by age

	Men	Women
Public: age≤30	114.78 (372)	113.46 (409)
age>30	155.73 (1883)	172.93 (1389)
Private formal: age≤30	117.17 (709)	108.55 (607)
age>30	156.05 (1127)	124.73 (596)
Private informal: age≤30	96.41 (1274)	61.01 (465)
age>30	109.30 (2120)	55.31 (796)
Self-employment: age≤30	148.47 (637)	99.97 (216)
age>30	171.60 (3742)	103.36 (1460)
Agriculture: age≤30	111.48 (805)	72.76 (214)
age>30	104.06 (4796)	78.25 (1765)

The number of observations is in parentheses; the earnings are estimated in Sri Lankan rupees (LKR); 1 USD ≈ 146 LKR

Next, we look at the predicted earnings by level of education for men and women. Table 9 presents the results. Education has a positive relationship with earnings – individuals with no education receive the lowest earnings while individuals with tertiary education receive the highest earnings. The increase in earnings by level of education is more pronounced in the public, formal private and self-employment sectors. Men with no education in the public sector earn 66 LKR per hour whereas men with tertiary education earn 213 LKR per hour, on average. For women, the increase in average earnings by level of education in the public sector is less pronounced, compared to men. The increase in earnings by level of education is greater for men working in agriculture, while there are slight differences across education levels in women's agricultural earnings. The informal sector and self-employment pay fairly similarly at lower levels of education, however stark differences are seen across the two sectors at the secondary and tertiary levels. Comparing the formal and informal private sectors, men with no education and tertiary education do comparatively better in terms of average earnings in the formal sector – men in other education levels (primary and secondary) benefit in the formal sector but the differences are less. Women in the formal sector with upper secondary or tertiary education earn relatively more in the formal sector compared to the informal sector. We found earlier in this section that individuals working in the public sector receive the highest wages. However a comparison between the public and formal private sectors of the average expected earnings by level of education suggests

that individuals with tertiary education are paid higher wages in the formal private sector, and not the public sector.

Table 9: Predicted (average) earnings per hour - by level of education

	Men	**Women**
Public:		
no education	65.83 (6)	49.57 (3)
primary	82.42 (55)	68.16 (22)
lower secondary	110.00 (184)	95.94 (24)
upper secondary	142.29 (1655)	143.60 (1155)
tertiary	213.43 (351)	196.80 (593)
Private formal:		
no education	40.15 (6)	47.64 (19)
primary	59.65 (58)	69.09 (64)
lower secondary	94.31 (162)	85.96 (123)
upper secondary	138.09 (1471)	118.45 (930)
tertiary	266.32 (138)	217.20 (64)
Private informal:		
no education	82.92 (38)	40.91 (42)
primary	90.94 (484)	46.92 (189)
lower secondary	101.70 (851)	54.13 (245)
upper secondary	108.92 (1990)	61.61 (775)
tertiary	131.57 (31)	80.30 (10)
Self-employment:		
no education	87.22 (44)	56.13 (40)
primary	110.97 (500)	69.10 (228)
lower secondary	145.31 (955)	86.16 (288)
upper secondary	184.03(2793)	113.44 (1082)
tertiary	294.61 (81)	185.79 (37)
Agriculture:		
no education	83.56 (228)	73.38 (219)
primary	92.75 (1759)	75.78 (642)
lower secondary	107.00 (1671)	78.07 (441)
upper secondary	116.96 (1908)	80.55 (668)
tertiary	136.27 (32)	81.19 (9)

The number of observations is in parentheses; the earnings are estimated in Sri Lankan rupees (LKR); 1 USD ≈ 146 LKR

6 Conclusion

This chapter analysed the determinants of employment choice and estimated the earnings in different employment sectors in Sri Lanka. The general literature on this subject in the context of Sri Lanka is inconclusive on some vital issues. Adding to this literature on labour market outcomes, firstly this chapter addressed the potential issues of sample selectivity and endogeneity of education. Secondly, it was able to separate the informal and formal private sectors for the wage employed, and the self-employed. As the informal sector in this country comprises of over 60 per cent of total employment, analysing this sector in isolation was necessary. Finally, this chapter attempted to establish differences in sector choice and earnings by gender.

Sample selectivity is a form of bias that arises from the non-random nature in which an individual may select into a particular sector of employment. As we do not have the wages of those who are not in the labour force, ignoring such individuals could lead to biased estimates. In order to deal with the potential issue of sample selectivity and examine the contributing factors towards choice of employment sector, Lee's multinomial logit (MNL) estimation was employed. Five sectors were identified – the non-agriculture sectors were separated into public, private wage (formal and informal) and self-employment, while agriculture employment combined the wage and self-employment in agricultural activities. The unemployed/not employed were also included in this part of the analysis, allowing us to control for sample selectivity. The predicted probabilities are obtained from the MNL and are included in the Mincerian earnings functions for individuals in various sectors of employment.

Furthermore, following studies by Garen (1984), Wooldridge (2005), Söderbom et al. (2006) and Kuépié et al. (2009), we use a Control Function (CF) approach to deal with the issue of endogeneity of education. Chapter 2 highlighted the importance of this form of bias which leads to downward-biased OLS estimates. The CF approach regressed education on all exogenous variables, the residuals were obtained, and were included in the MNL and earnings functions. We found that endogeneity of education is a bigger issue than sample selectivity in the case of Sri Lanka. Had we not

corrected for endogeneity of education, OLS estimates would be downward biased. The selectivity bias is primarily an issue for men in the formal private sector.

The findings suggest that the characteristics affecting sector choice and earnings vary greatly across sectors, and across men and women thus validating the necessity to analyse the sectors in isolation. The results from the MNL model of labour force participation suggest that more educated men get in to the public or formal private sector, while the least educated men get in to agriculture or the informal private sector. The most educated women are in the public sector. There were certain characteristics that vary by gender – married men are more likely to join the labour force, whereas married women are less likely to do so. The presence of young children or children of school age has no impact on men's decisions to participate in the labour force (with the exception of the formal private sector). For women however, the presence of children (young and of school-age) has an adverse impact on their choice to work.

The results for the earnings functions also show differences across sectors. The returns to education are highest in the formal private sector for both sexes – an extra year of education yields a 16 per cent increase in earnings, on average. Women receive similar returns in the public sector as well, whereas men receive considerably lower returns in the public sector in comparison to the formal private sector. The returns to education are lowest in the informal sector for both sexes, thus strengthening the argument for separating the formal and informal private sectors. The OLS estimates were helpful in predicting how much an individual earns in any particular sector after controlling for observable characteristics such as age, education, ethnicity and marital status. Earnings have a positive relationship with education and age. On average, women earn more than men only in the public sector. The literature suggests that the earnings tend to be fairly equal across men and women in the public sector, compared to the private sector. Appleton et al. (1999) and Gunatilaka (2008) and Tiagi (2010) obtained similar results where women in the public sector receive earnings similar to men, or even more. Overall, our findings are in line with those of Gunatilaka (2008) who identified the importance of disaggregating the private sector into the formal and informal sectors - characteristics, determination of sector and earnings vary across these two sectors. Additionally, we were able to extend the

analysis to include self-employed individuals and deal with the issue of endogeneity of education.

The topic of sectoral choice is multi-dimensional. In this study, we tried to uncover several dimensions such as sector choice and gender biases. However, it is important to note that 92 per cent of women in this sector are in white-collar jobs (41 per cent of formal private sector), in comparison to 65 per cent of men in the public sector in white-collar jobs (57 per cent of the formal private sector). Controlling for occupation/skill level in the analysis can be a further dimension for future research. The study employed the multinomial logit (MNL) which restricts us from doing so – the MNL examines the probabilities of choosing a particular sector and therefore it seems improper to include controls for occupation since sector choice and occupation are likely to go hand-in-hand, rather than a sequential process. Given that our results suggest education is an important factor in getting into high-paying jobs, a potential area for further research would be to explore the reasons as to why certain people get less education than others – possible explanations may include such individuals being constrained by poverty and/or doing relatively less well in school.

A limitation of the chapter was the focus on the mean, instead of exploring the entire earnings distribution. This was the consequence of the use of methodology; accounting for sample selection and endogeneity is more appropriate in the setting of MNL and OLS estimation. Despite this limitation, the study of sector choice adds to the existing theoretical and policy debates on the labour market by controlling for two forms of bias simultaneously (endogeneity and sample selectivity) and including informal sector employees and the self-employed in the analysis.

7 References

Adamchik, V. and Bedi, A., 2000, "Wage Differentials between the Public and Private Sectors: Evidence from an Economy in Transition", *Labour Economics*, 7 (2), pp. 203 – 224

Ajwad, M.J. and Kurukuasuriya, P., 2002, "Ethnic and Gender Wage Disparities in Sri Lanka", *Sri Lanka Economic Journal*, 3 (1), pp. 1 – 26

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

Bourguignon, F., Fournier, M., Gurgand, M., 2007, "Selection Bias Corrections Based on the Multinomial Logit Model: Monte-Carlo Comparisons", *Journal of Economic Surveys*, 21 (1), pp. 174 – 205

Christopoulou, R. and Monastiriotis, V., 2014, "The Greek Public Sector Wage Premium before the Crisis: Size, Selection and Relative Valuation of Characteristics", *British Journal of Industrial Relations*, 52 (3), pp. 579 – 602

Cotton, J., 1988, "On the decomposition of wage differentials", *Review of Economics and Statistics*, 70 (2), pp. 236 – 243

Dasgupta, S., 2003, "Structural and Behavioural Characteristics of Informal Service Employment: Evidence from a Survey in New Delhi", *the Journal of Development Studies*, 39 (3), pp. 51 - 80

Department of Census and Statistics, 2013, "Labour Force Survey 2013: Annual Report", Colombo, Sri Lanka

Fortin, M.N., Lemieux, T. and Firpo, S., 2010, "Decomposition methods in economics", In: Ashenfelter, O., Card, D. (Eds.), *Handbook of Labor Economics*, Volume 4, Part A, pp. 1 – 102. North Holland: Amsterdam

Frankema, E., 2006, "The Colonial Origins of Inequality: The Causes and Consequences of Land Distribution", *unpublished manuscript,* Groningen Growth and Development Centre, University of Groningen

Garen, J., 1984, "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable", *Econometrica*, 52 (5), pp. 1199 – 1218

Gardeazabal, J., and Ugidos, A., 2004, "More on identification in detailed wage decompositions", *Review of Economics and Statistics*, 86 (4), pp. 1034 – 1036

Gylfason, T. and Zoega, G., 2002, "Inequality and Economic Growth: Do Natural Resources Matter?" CESifo Working Paper Series No. 712

Gunatilaka, R., 2008, "Informal Employment in Sri Lanka: Nature, Probability of Employment, and Determinants of Wages", International Labour Organization, Colombo

Gunderson, M., 1979, "Earnings Differentials between the Public and Private Sectors", *Canadian Journal of Economics*, 12 (2), pp. 228 – 242

Gunerwardena, D., Abeyrathna, D., Ellagala, A., Rajakaruna, K. and Rajendran, S., 2008, "Glass Ceiling, Sticky Floors or Sticky Doors? A Quantile Regression Approach to Exploring Gender Wage Gaps in Sri Lanka", Poverty and Economic Policy Research Network, PMMA Working Paper Number 4

Hamilton, B., 2000, "Does Entrepreneurship Pay? An Empirical Analysis of the Returns to Self-Employment", *Journal of Political Economy*, 108 (3), pp. 604 – 631

Heckman, J., 1979, "Sample Selection Bias as a Specification Error", *Econometrica*, 47 (1), pp. 153 – 161

Heitmueller, A., 2004, "Public-Private Sector Wage Differentials in Scotland: An Endogenous Switching Model", the Institute for the Study of Labour, Discussion Paper No. 992

Hess, P. and Ross, C., 1997, "Economic Development: Theories", Dryden Press/Harcourt Brace Publishers, Evidence and Policies, Fortworth

Hyder, A. and Reilly, B., 2005, "The Public and Private Sector Pay Gap in Pakistan: A Quantile Regression Analysis", *the Pakistan Development Review*, 44 (3), pp. 271 – 306

Jaan, B., 2008, "The Blinder-Oaxaca decomposition for linear regression models", the Stata Journal, 8 (4), pp. 453 - 479

Kuépié, M., Nordman, C. and Roubaud, F., 2009, "Education and Earnings in Urban West Africa", *Journal of Comparative Economics*, 37, pp. 491 – 515

Department of Census and Statistics, 2011, "Labour Force Survey Report
Lee, L.-F., 1983, "Generalized Econometric Models with Selectivity", *Econometrica*, 51 (2), pp. 507 - 512

Mengistae, T., 1999, "Wage rates and job queues: Does the public sector overpay in Ethiopia?" Policy Research Working Paper Series, WPS-2105, the World Bank

Mulligan, C. and Rubinstein, Y., 2008, "Selection, Investment, and Women's Relative Wages Over Time", *Quarterly Journal of Economics*, 123 (3), pp. 1061 – 1110

Neumark, D., 1988, "Employers' discriminatory behaviour and the estimation of wage discrimination", *Journal of Human Resources*, 23 (3), pp. 279 – 295

Oaxaca, R. and Ransom, M., 1994, "On discrimination and the decomposition of wage differentials", *Journal of Econometrics*, 61 (1), pp. 5 – 21

Oaxaca, R. and Ransom, M., 1999, "Identification in detailed wage decompositions", *Review of Economics and Statistics*, 81 (1), pp. 154 – 157

Parker, S., 2009, "The Economics of Self-Employment and Entrepreneurship," Cambridge, Cambridge University Press

Pratap, S. and Quintin, E., 2006, "Are labor markets segmented in developing countries? A semiparametric approach", *European Economic Review*, 50 (7), pp. 1817 – 1841

Ray, D., 1998, "Development Economics", Princeton University Press, New Jersey

Reimers, C., 1983, "Labor market discrimination against Hispanic and black men", *Review of Economics and Statistics*, 65 (4), pp. 570 – 579

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 (Supplement 2), ii95 – ii148

Schwiebert, J., 2015, "Estimation and Interpretation of a Heckman Selection Model with Endogenous Covariates", *Empirical Economics*, 49 (2), pp. 675 – 703

Söderbom, M., Teal, F., Wambugu, A., Kahyarara, G., 2006, "Dynamics of Returns to Education in Kenyan and Tanzanian Manufacturing", *Oxford Bulletin of Economics and Statistics*, 68 (3), pp. 261 – 288

Tansel, A., 2004, "Public-Private Employment Choice, Wage Differentials and Gender in Turkey", the Institute for the Study of Labour, Discussion Paper No. 1262

Terrell, K., 1993, "Public-private wage differentials in Haiti: Do public servants earn a rent?" *Journal of Development Economics*, 42 (2), pp. 293 - 314

Tiagi, R., 2010, "Public Sector Wage Premium in Canada: Evidence from a Labour Force Survey", *LABOUR*, 24 (4), pp. 456 – 473

Yun, M.-S., 2005, "Hypothesis tests when decomposing differences in the first moment", *Journal of Economic and Social Measurement*, 30 (4), pp. 295 - 304

Vijverberg, W.P., 1995, "Returns to Schooling in Non-Farm Self-Employment: An Econometric Case Study of Ghana", *World Development*, 23 (7), 1215 – 1227

Wooldridge, J., 2005, “Unobserved Heterogeneity and Estimation of Average Partial Effects”, In *Identification and Inference for Econometric Models,* prepared under the direction of D. W. K. Andrews and J. H. Stock, pp. 27 – 55. Cambridge U.K.: Cambridge University Press

Wooldridge, J., 2010, “Econometric Analysis of Cross Section and Panel Data”, 2nd Edition, the MIT Press, Cambridge

Appendix 1

Table A1: Descriptive statistics (averages) for the full sample

Variables	Non-agriculture				Agriculture	Not in labour force
	Public	Private wage (formal)	Private wage (informal)	Self-employed		
Earnings per hour	134.68	112.54	68.43	109.10	68.97	
Hours worked per week	44.8	49.8	46.4	45.1	38.2	
Years of age	40.7	34.5	37.8	44.5	46.6	31.7
Urban	0.21	0.24	0.19	0.22	0.03	0.17
Sinhalese	0.81	0.85	0.73	0.75	0.70	0.69
Married	0.82	0.63	0.64	0.82	0.82	0.38
Female	0.44	0.40	0.27	0.28	0.26	0.66
Years in education	12.64	11.04	8.99	9.38	6.93	7.38
Industry(base: services)						
Manufacturing	0.05	0.54	0.46	0.33		
Occupation (base: other)						
High skilled white collar	0.53	0.31	0.08	0.21	0.01	
Low skilled white collar	0.24	0.19	0.20	0.22		
High skilled blue collar	0.03	0.21	0.30	0.34	0.57	
Low skilled blue collar	0.16	0.29	0.42	0.22	0.42	
Number of observations	4,053	3,039	4,655	6,055	7,580	43,871

Table A2: Reduced form for education: full sample

Dependent variable: years in education		
Variables	**Coefficient**	**z-statistic**
Policy	3.499***	66.95
Spouse's education	0.566***	110.05
Mother's education	0.114***	23.01
Father's education	0.028***	5.55
Spouse: in public sector	0.287***	5.45
in private sector	-0.059*	-1.65
Mother: in public sector	-0.064	-1.41
in private sector	0.029	0.92
Father: in public sector	-0.164***	-3.57
in private sector	0.036	1.04
Proportion of children (age<6)	-0.995***	-8.63
Proportion of children (6≤age≤18)	-1.415***	-18.04
Proportion of elders (age>65)	1.702***	16.96
Age	1.148***	231.56
Age^2	-0.022***	-140.66
Age^3	$1.28x10^{-4}$***	91.16
Urban	0.711***	25.88
Sinhalese	0.697***	28.86
Married	-4.613***	-71.11
Female	0.117***	5.48
Constant	-9.620***	-123.05

N=69,253; F(20, 69229)=8528.16; Prob>F=0.00; R^2=0.57

F-test to check for significance of instruments in determining education: F(13, 69229)=1317.25; Prob>F=0.00

Controlled for missing values of spouse and parent(s) employment sector and/or level of education; * p<0.1, ** p<0.05, ***p<0.01; Default categories are rural, non-Sinhalese, unmarried, male, spouse/mother/father being unemployed/not employed.

Table A3: Determinants of probability of employment (base category: unemployed/not employed) – full sample

Variables	Non-agriculture								Agriculture	
	Public		Private wage (formal)		Private wage (informal)		Self-employed			
	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic
Individual characteristics:										
Education (years)	0.367***	14.69	0.102***	3.73	-0.174***	-9.06	-0.034**	-1.97	-0.287***	-19.29
Residuals from first stage	0.095***	3.58	0.058**	2.00	0.155***	7.75	0.069***	3.78	0.178***	11.52
Age	0.802***	11.05	1.149***	21.78	1.111***	32.84	1.114***	28.13	1.027***	31.56
Age^2	-0.010***	-6.20	-0.022***	-18.60	-0.021***	-29.57	-0.198***	-25.32	-0.018***	-27.06
Age^3	$1.33x10^{-5}$	1.08	$1.14x10^{-4}$***	13.81	$1.11x10^{-4}$***	24.38	$1.04x10^{-4}$***	21.31	$8.69x10^{-5}$***	20.82
Urban	-0.213***	-3.76	0.302***	5.38	0.187***	3.86	0.127***	2.89	-1.790***	-24.23
Sinhalese	0.356***	6.35	0.930***	14.93	0.477***	10.77	0.306***	7.39	0.369***	9.61
Married	-0.275***	-3.23	-0.461***	-5.40	-0.422***	-5.94	-0.341***	-4.99	-0.092	-1.47
Female	-1.905***	-40.21	-1.928***	-40.16	-2.417***	-57.55	-2.397***	-61.42	-2.583***	-69.33
Policy (=1 if age $\leq$ 26)	0.797***	5.31	1.240***	9.33	1.618***	15.52	0.887***	7.47	1.570***	14.38
Education:										
Spouse	0.073***	7.54	0.068***	6.60	0.054***	6.61	0.076***	10.13	0.042***	6.07
Mother	-0.023*	-1.89	0.010	0.89	-0.045***	-4.75	-0.025**	-2.29	-0.096***	-8.58
Father	-0.046***	-3.27	-0.011	-0.93	-0.062***	-5.50	-0.030**	-2.29	0.128***	-8.72
Sector of employment:										
Spouse: in public sector	0.309***	4.34	-0.497***	-5.13	-1.183***	-10.20	-0.747***	-9.98	-0.157**	-2.17
in private sector	-0.808***	-13.12	-0.509***	-7.94	-0.708***	-13.21	-0.805***	-17.27	-0.909***	-20.25
Mother: in public sector	-0.139	-0.68	-0.602***	-3.28	-0.851***	-3.88	-0.614**	-2.31	0.360*	1.67
in private sector	0.120	1.06	0.108	1.23	0.168**	2.17	0.113	1.11	0.128	1.27
Father: in public sector	0.258	1.47	-0.602***	-3.88	-0.509***	-3.09	-0.672***	-3.00	-0.562**	-2.41

in private sector	-0.061	-0.64	-0.347***	-4.38	-0.020	-0.27	-0.128	-1.44	-0.385***	-4.05
Household characteristics:										
Proportion of children (age<6)	-0.068	-0.37	-1.591***	-7.70	-1.006***	-5.86	0.044	0.29	-0.111	-0.77
Proportion of children (6≤age≤18)	-0.172	-1.37	-1.137***	-7.74	-0.713***	-6.26	0.161	1.64	0.039	0.42
Proportion of elders (age>65)	0.015	0.06	0.374	1.41	0.427***	2.60	0.600***	4.97	1.032***	9.92
Constant	-19.477***	-20.04	-19.828***	-30.00	-15.288***	-35.44	-17.731***	-30.60	-13.203***	-28.52
N = 69, 253; LR chi^2(120) = 53689.03; Prob>chi^2 = 0.00; Pseudo R^2 = 0.32										

Controlled for missing values of spouse and parent(s) employment sector and/or level of education; * p<0.1, ** p<0.05, ***p<0.01; Default categories are rural, non-Sinhalese, unmarried, male, spouse/mother/father being unemployed/not employed.

Table A4: Earnings functions – full sample

Dependent variable: log(earnings per hour)	Non-agriculture								Agriculture	
	Public		**Private wage (formal)**		**Private wage (informal)**		**Self-employed**			
	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic
Education (years)	0.102***	6.19	0.170***	14.88	0.047***	4.25	0.106***	10.36	0.045***	4.06
Age	0.108*	1.76	0.090**	2.42	0.111***	4.94	0.127***	3.12	0.051	1.64
Age^2	-0.002	-1.20	-0.001	-1.56	-0.002***	-4.40	-0.002**	-2.23	-0.001	-1.53
Age^3	$9.49x10^{-6}$	0.87	$5.53x10^{-6}$	0.87	$1.29x10^{-5}$***	3.70	$7.56x10^{-6}$	1.31	$6.20x10^{-6}$	1.34
Urban	0.058	1.52	0.220***	5.41	0.221***	8.00	0.288***	6.24	0.202	1.61
Sinhalese	0.009	0.25	-0.130***	-2.70	-0.065**	-2.07	-0.163***	-4.42	-0.177***	-6.06
Married	-0.028	-0.67	0.038	0.71	0.053	1.36	0.068	1.29	0.105**	2.53
Female	-0.005	-0.17	-0.291***	-7.49	-0.539***	-8.94	-0.776***	-11.02	-0.226***	-3.15
Residuals from first stage	-0.033***	-3.66	-0.072***	-6.72	-0.023*	-1.92	-0.045***	-4.42	-0.021**	-2.44
Lambda (selectivity correction)	-0.016	-0.24	0.242***	2.75	0.195**	2.23	0.341***	3.36	-0.028	-0.29
Constant	1.390	1.47	0.647	1.00	1.722***	4.43	0.472	0.66	2.874***	6.17
Number of observations	4,053		3,039		4,655		6,055		7,580	
R^2	0.14		0.17		0.07		0.10		0.02	

* p<0.1, ** p<0.05, ***p<0.01; Default categories are rural, non-Sinhalese, unmarried, male: the standard errors are bootstrapped with 500 replications

Table A5: Reduced form for education: men

Dependent variable: years in education		
Variables	**Coefficient**	**z-statistic**
Policy	3.701***	45.65
Spouse's education	0.585***	79.93
Mother's education	0.123***	18.80
Father's education	0.034***	4.97
Spouse: in public sector	0.471***	5.87
in private sector	-0.290***	-4.97
Mother: in public sector	-0.056	-0.91
in private sector	0.096**	2.20
Father: in public sector	-0.161**	-2.53
in private sector	0.028	0.61
Proportion of children (age<6)	-1.311***	-7.47
Proportion of children (6≤age≤18)	-1.680***	-14.58
Proportion of elders (age>65)	1.554***	9.71
Age	1.119***	156.79
Age^2	-0.021***	-89.90
Age^3	$1.21x10^{-4}$***	56.65
Urban	0.752***	19.21
Sinhalese	0.237***	7.01
Married	-5.267***	-51.24
Constant	-9.562***	-84.08

N=32,275; F(19, 32252)=4346.84; Prob>F=0.00; R^2=0.59

F-test to check for significance of instruments in determining education: F(13, 32252)=382.95; Prob>F=0.00

Controlled for missing values of spouse and parent(s) employment sector and/or level of education; * p<0.1, ** p<0.05, ***p<0.01; Default categories are rural, non-Sinhalese, unmarried, spouse/mother/father being unemployed/not employed.

Table A6: Reduced form for education: women

Dependent variable: years in education		
Variables	**Coefficient**	**z-statistic**
Policy	3.009***	44.76
Spouse's education	0.579***	80.72
Mother's education	0.101***	13.47
Father's education	0.021***	2.89
Spouse: in public sector	-0.111	-1.48
in private sector	-0.168***	-3.11
Mother: in public sector	-0.079	-1.18
in private sector	0.034	-0.76
Father: in public sector	-0.174***	-2.67
in private sector	0.021	0.43
Proportion of children (age<6)	-0.636***	-4.27
Proportion of children (6≤age≤18)	-1.277***	-12.10
Proportion of elders (age>65)	1.618***	12.56
Age	1.180***	169.44
Age^2	-0.023***	-107.18
Age^3	$1.33x10^{-4}$***	70.87
Urban	0.646***	17.08
Sinhalese	1.082***	32.50
Married	-4.616***	-52.34
Constant	-9.310***	-84.33
N=36,978; F(19, 36955)=4771.83; Prob>F=0.00; R^2=0.57		
F-test to check for significance of instruments in determining education: F(13, 32252)=689.84; Prob>F=0.00		

Controlled for missing values of spouse and parent(s) employment sector and/or level of education; * p<0.1, ** p<0.05, ***p<0.01; Default categories are rural, non-Sinhalese, unmarried, male, spouse/mother/father being unemployed/not employed.

Table A7: Earnings functions with controls for occupation type - men

Dependent variable: log(earnings per hour)	Non-agriculture								Agriculture	
	Public		Private wage (formal)		Private wage (informal)		Self-employed			
	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic
Education (without occupation dummies)	**0.088*** **	**4.69**	**0.162*** **	**11.94**	**0.043*** **	**3.48**	**0.093*** **	**8.02**	**0.054*** **	**3.97**
Education (years)	0.072***	3.70	0.122***	8.26	0.043***	3.47	0.086***	7.27	0.056***	4.10
Age	0.111	1.51	0.027	0.67	0.118***	4.52	0.088*	1.78	0.054*	1.66
Age^2	-0.001	-1.06	$6.28x10^{-5}$	0.07	-0.002***	-4.08	-0.001	-1.24	-0.001	-1.62
Age^3	$9.99x10^{-6}$	0.75	$-5.57x10^{-6}$	-0.77	$1.42x10^{-5}$***	3.50	$4.21x10^{-6}$	0.60	$7.59x10^{-6}$	1.46
Urban	0.087*	1.80	0.204***	4.43	0.173***	5.44	0.258***	4.73	0.217	1.63
Sinhalese	-0.011	-0.28	-0.050	-1.07	-0.085**	-2.46	-0.135***	-3.39	-0.219***	-6.95
Married	-0.051	-0.99	0.208***	3.18	0.150***	3.00	0.141*	1.87	0.191***	3.27
Residuals from first stage	-0.019	-1.47	-0.034***	-2.69	-0.016	-1.26	-0.028**	-2.46	-0.022**	-2.23
Lambda (selectivity correction)	0.005	0.06	0.179**	2.37	0.012	0.11	0.117	0.96	-0.068	-0.55
High skilled white collar	-0.065	-1.08	0.470*	1.71	-0.216	-0.71	0.169	0.91	-0.180	-0.97
Low skilled white collar	-0.280***	-4.58	0.113	0.41	-0.482	-1.62	-0.081	-0.43	-0.045	-0.19
High skilled blue collar	-0.074	-1.13	0.132	0.48	-0.275	-0.93	0.040	0.22	-0.838***	-26.64
Low skilled blue collar	-0.304***	-5.44	0.094	0.34	-0.384	-1.30	-0.029	-0.16	-0.721***	-19.23
Constant	1.931**	1.96	1.789**	2.42	2.255***	4.30	1.662*	1.88	3.615***	7.48
Number of observations	2,255		1,836		3,394		4,379		5,601	
R^2	0.13		0.21		0.04		0.06		0.03	

* p<0.1, ** p<0.05, ***p<0.01; Default categories are rural, non-Sinhalese, unmarried, other "collar" employment (that is, Armed Forces occupations and workers not classified by occupations); the standard errors are bootstrapped with 500 replications

Table A8: Earnings functions with controls for occupation – women

Dependent variable: log(earnings per hour)	Non-agriculture								Agriculture	
	Public		Private wage (formal)		Private wage (informal)		Self-employed			
	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic
Education (without occupation dummies)	**0.141*** **	**5.40**	**0.168*** **	**9.02**	**0.058** **	**2.51**	**0.126*** **	**6.23**	**0.014**	**0.49**
Education (years)	0.126***	4.92	0.114***	5.59	0.050**	2.31	0.119***	5.70	0.030	1.02
Age	0.120	0.96	0.018	0.22	0.010	0.24	0.107	1.21	0.038	0.39
Age^2	-0.002	-0.59	$-9.43x10^{-6}$	-0.01	$-1.02x10^{-4}$	-0.10	-0.002	-0.91	$-6.93x10^{-4}$	-0.37
Age^3	$9.44x10^{-6}$	0.39	$-1.33x10^{-6}$	-0.10	$-2.48x10^{-7}$	-0.04	$6.99x10^{-6}$	0.66	$4.368x10^{-6}$	0.38
Urban	0.002	0.04	0.149**	2.27	0.292***	5.84	0.289***	3.06	-0.107	-0.20
Sinhalese	0.047	0.78	-0.202	-1.51	0.014	0.15	-0.141	-1.44	0.072	1.07
Married	-0.016	-0.27	0.015	0.12	-0.047	-0.67	$4.17x10^{-4}$	0.00	0.080	1.05
Residuals from first stage	-0.046***	-3.97	-0.072***	-4.07	-0.021	-0.89	-0.063***	-2.94	-0.016	-0.76
Lambda (selectivity correction)	0.084	0.89	0.067	0.26	0.058	0.43	0.323	1.34	-0.058	-0.27
High skilled white collar	0.199	1.18	0.204	0.59	-0.250**	-2.40	-0.696***	-4.89	0.535	0.70
Low skilled white collar	0.022	0.13	-0.095	-0.27	-0.392***	-5.11	-0.773***	-6.07		
High skilled blue collar	-0.429	-1.39	-0.271	-0.77	-0.536***	-5.86	-0.998***	-8.01	-0.194	-0.26
Low skilled blue collar	-0.129	-0.70	-0.77	-0.79	-0.299***	-3.03	-0.768***	-4.91	0.030	0.04
Constant	0.442	0.24	2.591	1.64	3.191***	4.43	0.767	0.43	2.777	1.54
Number of observations	1,798		1,203		1,261		1,676		1,979	
R^2	0.19		0.14		0.06		0.06		0.01	

* p<0.1, ** p<0.05, ***p<0.01; Default categories are rural, non-Sinhalese, unmarried, other "collar" employment (that is, Armed Forces occupations and workers not classified by occupations); the standard errors are bootstrapped with 500 replications

Table A9: Predicted earnings: full sample

	Non-agriculture				Agriculture
	Public	**Private wage (formal)**	**Private wage (informal)**	**Self-employment**	
Predicted earnings	153.59 (4053)	132.54 (3039)	90.51 (4655)	151.91 (6055)	97.72 (7580)
By age:					
≤ 30	114.68	112.82	86.76 (1739)	138.14 (853)	102.06 (1019)
> 30	(781)	(1316)	92.74 (2916)	154.17 (5202)	97.05 (6561)
	162.88	147.60			
	(3272)	(1723)			
By level of education:					
No education	59.17 (9)	40.02 (25)	61.79 (80)	71.12 (84)	78.02 (447)
Primary	77.56 (77)	62.09 (122)	77.03 (673)	98.73 (728)	88.93 (2401)
Lower secondary	106.06	91.89 (285)	89.03 (1096)	133.69 (1243)	100.47 (2112)
Upper secondary	(208)	132.05	94.78 (2765)	166.16 (3875)	106.68 (2576)
Tertiary	142.94	(2401)	119.58 (41)	263.53 (118)	122.51 (41)
	(2810)	250.86			
	203.23	(202)			
	(944)				

The number of observations is in parentheses; the earnings are estimated in Sri Lankan rupees (LKR); 1 USD ≈ 146 LKR

Table A10: Variables used for the analysis

Variable	Description	Number of observations	Mean	Minimum	Maximum	Standard deviation
log(earned income)	natural log of earned income	25,382	4.15	0	9.76	1.17
education (years)	total years of education	25,382	9.30	0	19	3.74
age	individual's age	25,382	42.12	15	65	13.47
female	female=1 if the individual is a woman, else zero (man)	25,382	0.31	0	1	0.46
urban	urban=1 if the individual is from an urban area, else zero (rural)	25,382	0.16	0	1	0.40
married	married=1 if the individual is reported to be married, else zero (unmarried/divorced/widowed/separated)	25,382	0.76	0	1	0.43
Sinhalese	Sinhalese=1 if the individual is from a Sinhalese ethnic background, else zero (Muslim/Tamil/Burgher/Other)	25,382	0.75	0	1	0.43
policy	change in compulsory schooling to 14 years in 1997 (policy=1 if the individual is in school from 1997 onwards, else zero)	25,382	0.14	0	1	0.34
spouse's education	years of education obtained by spouse, if the individual is married	23,641	6.91	0	19	5.13
mother's education	years of education obtained by the mother, if the information is available	6,349	7.01	0	19	3.92
father's education	years of education obtained by the father, if the information is available	4,476	7.33	0	19	3.77
spouse-public sector	=1 if spouse is in the public sector, else zero	23,641	0.08	0	1	0.28
spouse-private sector	=1 if spouse is in the private sector, else zero	23,641	0.20	0	1	0.40
spouse-unemployed/not employed	=1 if spouse is unemployed/not employed, else zero	23,641	0.72	0	1	0.45
mother-public sector	=1 if mother is in the public sector, else zero	6,349	0.02	0	1	0.15
mother-private sector	=1 if mother is in the private sector, else zero	6,349	0.16	0	1	0.36

mother-unemployed/not employed	=1 if mother is unemployed/not employed, else zero	6,349	0.82	0	1	0.38
father-public sector	=1 if father is in the public sector, else zero	4,476	0.05	0	1	0.22
father-private sector	=1 if father is in the private sector, else zero	4,476	0.51	0	1	0.50
father-unemployed/not employed	=1 if father is unemployed/not employed, else zero	4,476	0.44	0	1	0.50
proportion of children (age<6)	the proportion of children in the household below the age of 6	25,382	0.06	0	0.67	0.12
proportion of children (6≤age≤18)	the proportion of children in the household between the ages of 6 and 18	25,382	0.15	0	0.80	0.20
proportion of elders (age>65)	the proportion of individuals in the household over the age of 65	25,382	0.04	0	0.75	0.14
high-skilled white collar	=1 is individual is a high-skilled white collar worker	25,382	0.19	0	1	0.39
low-skilled white collar	=1 is individual is a low-skilled white collar worker	25,382	0.15	0	1	0.36
high-skilled blue collar	=1 is individual is a high-skilled blue collar worker	25,382	0.34	0	1	0.47
low-skilled blue collar	=1 is individual is a low-skilled blue collar worker	25,382	0.31	0	1	0.46
other occupation	=1 is individual is in the armed forces or unclassified occupation	25,382	0.01	0	1	0.08

Appendix 2: Blinder-Oaxaca decomposition

Here, we examine the average wage gap between men and women using the results of the estimated earnings functions that account for selectivity and endogeneity of education. For the purpose of the decomposition, the full sample estimates were used. The Blinder-Oaxaca (1973) technique is employed to decompose the wage gap. While chapter 3 utilizes the three-fold decomposition for the quantile regression, in this chapter we focus on the two-fold decomposition at the mean. The technique is explained below.

The decomposition enables us to identify how much of the mean earnings differential,

$$R = E(Y_M) - E(Y_W)$$

where E(Y) is the expected value of earnings for men and women, is due to differences in male/female characteristics (for example, education, age, ethnicity).

The overall mean differential between men and women can be rearranged in the following way:

$$R = \{E(X_M) - E(X_W)\}'\beta^* + \{E(X_M)'(\beta_M - \beta^*) + E(X_W)'(\beta^* - \beta_W)\}$$

where β^* is a non-discriminatory coefficient vector[61].

There are two components in the above equation: $R = Q + U$

[61] There are several ways in which β^* is measured. One way to do this is to assume that there is discrimination directed towards one group only – men or women. In this case, $\beta^* = \beta_M$ or $\beta^* = \beta_W$. However, there is no particular reason to assume that there is no discrimination towards either gender. Other studies have chosen to use the average coefficients over both groups (Reimers, 1983) or weight the coefficients by the respective sizes of each group (Cotton, 1988) to obtain an estimate of the non-discriminatory coefficient vector. Neumark (1988) and Oaxaca and Ransom (1994) use the coefficients from a pooled regression over both groups (men and women) as an estimate for β^*; relative weights are given to the coefficients of each group.

The first component is: $Q = \{E(X_M) - E(X_W)\}'\beta^*$

This is the part of the earnings differential explained by differences in the characteristics between men and women, known as the "endowments effect".

The second component is: $U = E(X_M)'(\beta_M - \beta^*) + E(X_W)'(\beta^* - \beta_W)$

This is the unexplained part (known as the "discrimination effect"), which is usually attributed to discrimination. Additionally, it captures the potential effects of differences in unobservable characteristics across men and women.

The use of categorical variables could give rise to a potential problem with the decomposition results since the results for the categorical variables depend on the choice of omitted category (Oaxaca and Ransom, 1999; Gardeazabal and Ugidos, 2004; Yun, 2005). Categorical variables such as marital status, urban/rural and ethnicity are measured as dummy variables (with a value of zero or one), and one category is omitted due to collinearity (the base category). Jaan (2008) states that the magnitudes of the endowment and discrimination effects, and the contribution of a single indicator variable (that is, the total contribution of the categorical variable) towards the explained part are not affected by the choice of base category. However for the unexplained part, the choice of base category is critical – there is a trade-off in the unexplained part between the component that is attributed to differences in intercepts and the component attributable to differences in slope coefficients. Thus, a change in base category would result in a change in the results for a specific dummy variable as well as the contribution of the entire categorical variable. Yun (2005) suggested a solution to this problem - the model is estimated using dummy variable coding, then the coefficient vectors are transformed so that the deviations from the grand mean are expressed and the (redundant) coefficient for the base category is included (Yun, 2005; Jaan, 2008). This transformation is used in the decomposition[62].

[62] However, Fortin et al. (2010) state that there is no general solution to deal with the issue of categorical variables in decomposition techniques.

The results of the twofold Blinder-Oaxaca decomposition are presented as follows. The decomposition was estimated using the MNL and reduced form (for education) estimates for the full sample which included a female dummy. It is not possible to carry out the decomposition based on the MNL and reduced form estimates of the sub-samples by gender as the decomposition requires the full sample, which is then disaggregated by gender. The average log earnings per hour are highest in the public sector for both, men (4.68) and women (4.77) and lowest in agriculture employment. With the exception of the public sector, men receive higher average earnings compared to women in the other sectors of employment. The mean wage gap between men and women is highest for the self-employed (62 per cent), followed by the informal (46 per cent) and formal (36 per cent) private sectors. The wage gap is split into two elements – the first element ("explained" part) measures the part of the predicted wage gap due to differences in male and female characteristics. The second element ("unexplained" part) measures the possible discrimination towards either men or women, along with the differences in unobservable characteristics having an impact on the overall wage gap. In the public sector, the explained gap explains most of the wage gap. In the other sectors however, the unexplained gap explains a large proportion of the wage gap – especially in the informal sector and self-employment. In the formal sector and agriculture, the unexplained gap explains 81 and 83 per cent of the wage gap, respectively.

We move on to the discussion of the explained and unexplained elements in more detail for each sector. In the public sector, differences between male and female characteristics such as the years of education account for a substantial proportion of the wage gap leading to women earning higher wages than men. However in the formal private sector and agriculture, differences in education levels contribute towards a wage gap in favour of men. Being from urban areas, men benefit from higher wages in the formal sector. Additionally, selecting into the formal sector contributes towards a lower wage gap. In self-employment and the informal sector, the explained part is negative suggesting that average female characteristics are larger than average male characteristics - differences in characteristics such as being from urban areas and a Sinhalese background lead to higher wage gap, while differences in selection into this sector between the two genders leads to a lower gap.

The unexplained part of the wage gap is not significant at the 1 per cent level in the public sector. This part of the wage gap indicates how much of the total wage gap is contributable to differences across the sexes in unobservable characteristics and/or the presence of a female (male) disadvantage (advantage). In the formal private sector, the unexplained element explains 81 per cent of the total wage gap. In the informal sector and self-employment, most of the wage gap is due to the unexplained part of the decomposition – for instance in self-employment, it appears to be that differences in unobservable characteristics (as the variables used to estimate earnings do not have substantial coefficients, while the constant does) account for a large proportion of the wage gap.

Table A11: Blinder-Oaxaca decomposition

Dependent variable: log(earnings per hour)	Non-agriculture								Agriculture	
	Public		Private wage (formal)		Private wage (informal)		Self-employed			
	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic	Coefficient	z-statistic
Overall:										
Men ($\overline{Y_M}$)	4.680***	276.78	4.543***	216.53	4.110***	230.96	4.341***	229.13	3.902***	238.00
Women ($\overline{Y_W}$)	4.773***	261.80	4.183***	136.00	3.650***	145.27	3.718***	112.83	3.631***	134.78
Difference: ($\overline{Y_M} - \overline{Y_W}$)	-0.093***	-3.74	0.360***	9.66	0.461***	14.97	0.623***	16.38	0.271***	8.58
Explained:$\{E(X_M) - E(X_W)\}'\beta^*$	-0.098***	-5.10	0.069**	2.21	-0.078	-1.51	-0.154**	-2.57	0.044	0.68
Unexplained: $E(X_M)'(\beta_M - \beta^*) + E(X_W)'(\beta^* - \beta_W)$	0.005	0.17	0.291***	7.48	0.539***	8.94	0.776***	11.02	0.226***	3.15
Explained:										
Education (years)	-0.173***	-5.91	0.087***	4.65	-0.004	-0.74	0.011	1.04	0.029***	3.43
Age	0.216*	1.70	0.264**	2.30	-0.063	-1.17	-0.065	-1.28	0.019	0.93
Age^2	-0.284	-1.18	-0.283	-1.52	0.126	1.38	0.068	0.94	-0.079	-1.30
Age^3	0.102	0.86	0.067	0.87	-0.050	-1.33	-0.013	-0.58	0.051	1.24
Area (urban or rural)	$3.97X10^{-4}$	0.51	0.011***	2.86	$-4.64x10^{-4}$	-0.16	0.014***	3.45	0.006	1.58
Ethnicity	$-5.54X10^{-5}$	-0.22	0.004*	1.72	0.008**	2.02	0.012***	3.60	-0.005**	-2.13
Marital status	-0.003	-0.67	0.006	0.71	0.012	1.36	0.009	1.28	0.015**	2.49
Residuals from first stage	0.044***	3.57	-0.011	-1.63	0.002	0.93	0.014***	2.86	-0.008**	-2.14
Selectivity correction	-0.001	-0.24	-0.076***	-2.73	-0.109**	-2.23	-0.202***	-3.36	0.016	0.29
Unexplained:										

Education (years)	-1.031**	-2.53	-0.036	-0.09	-0.115	-0.49	-0.398*	-1.77	0.220	1.28
Age	-1.007	-0.17	0.539	0.14	3.758*	1.91	-3.915	-0.76	1.169	0.28
Age^2	-0.109	-0.02	-0.356	-0.12	-3.443*	-1.80	3.437	0.71	-1.234	-0.29
Age^3	0.277	0.13	-0.018	-0.02	1.049*	1.67	-1.074	-0.70	0.342	0.24
Area (urban or rural)	-0.034	-0.01	0.011	0.01	0.037	0.01	-0.014	-0.01	0.015	0.07
Ethnicity	-0.027	-0.01	-0.062	-0.01	-0.014	-0.01	0.004	-0.05	-0.051	-0.01
Marital status	-0.019	-0.03	0.017	0.17	-0.013	-0.24	0.035	0.03	0.032	0.30
Residuals from first stage	0.063*	1.87	0.014*	1.71	-0.003	-0.31	0.005	1.28	0.014	0.58
Selectivity correction	-0.293*	-1.67	0.353	0.53	0.053	0.15	-0.528	-0.95	0.125	0.35
Constant	2.184	1.04	-0.273	-0.11	-0.796	-0.92	3.196	1.37	-0.239	-0.15
Number of observations	4,053		3,039		4,655		6,055		7,580	

* p<0.1, ** p<0.05, ***p<0.01; the categorical variables (area, ethnicity and marital status) have been normalized; that is, the effects of the variables are interpreted as deviation contrasts from the grand mean as suggested by Yun (2005)

Chapter 5: Conclusion, summary of findings and scope for future research

This thesis explored the labour market outcomes in the context of Sri Lanka by examining three key issues: (1) investigating the effect of education on earnings and productivity, (2) recognising the contributors of the urban-rural welfare gap, and (3) evaluating the determinants of sector choice.

1 Summary of findings and contributions to literature

1.1 Rates of return to education in Sri Lanka

The first part of the thesis on the rates of returns to education in Sri Lanka dealt with the issue of endogeneity in education due to the omitted variable bias. It had three main objectives: (1) to estimate the rates of returns to education while dealing with endogeneity, (2) to estimate the returns at different levels of education (3) to analyse the effect of education on productivity of the self-employed. The data set used in this chapter was the Household Income and Expenditure Survey for 2009/10. The chapter contributes to the literature on rates of returns to education in Sri Lanka, a country that emphasises on the need to provide equal access to education. To our knowledge[63], there is no existing study that explores the issue of endogeneity of education in the context of this country.

For the first part of this chapter which dealt with endogeneity, an Instrumental Variables approach was used to estimate the earnings functions for individuals in agriculture, non-agriculture private and public sectors, and men and women separately. The Hausman test for exogeneity of education was rejected, implying that education is indeed endogenous. Two instruments were used to account for endogeneity of education - (1) a change in schooling laws that occurred in 1997, which made it compulsory for individuals to remain in school until the age of 14, and (2) spouse's education. After ensuring that the instruments are relevant (related to the

[63] Based on a literature search in EconLit

main variable of interest, education) and exogenous (have no partial effect on earnings after controlling for education), IV was chosen as the preferred method of analysis. The IV estimates of the returns to education were higher for all sub-samples compared to the OLS estimates. Ignoring the endogeneity of the education variable would have lead us to believe downward-biased OLS estimates. The findings suggest that women receive higher returns to education compared to men, even after controlling for the type of occupation and industry the individual is in.

Next, we estimated the earnings functions using levels of education (primary, lower secondary, upper secondary, and tertiary). For this, Ordinary Least Squares was used as the preferred method of estimation due to a lack of good instruments. This allowed us to identify whether the returns to education are higher at the tertiary level, or at the primary level. The results suggest that education and earnings share a convex relationship, that is, the returns to education are increasing with every level of education). This is in line with the findings by Colclough et al. (2010). We found that women had higher returns to education in comparison men at the upper secondary and tertiary levels; however at the primary and lower secondary levels of education, women had lower returns to education in comparison to men.

The final part of this chapter estimated a translog production function for the self-employed at the household level. This examined the link between education, farm/non-farm inputs (labour, raw materials and land) and productivity, rather than earnings. The results suggest that agricultural productivity is negatively associated with the presence of individuals with higher education; whereas non-agricultural output is positively associated with the presence of more educated individuals.

1.2 Decomposing the urban-rural welfare gap in Sri Lanka

To explore the recent developments in the country, the second part of the thesis estimated the urban-rural welfare gap for Sri Lanka in 2002 and 2009/10 using the Household Income and Expenditure Surveys for the respective years. The chapter had two main objectives: (1) to identify the household-level determinants of per capita expenditure across urban and rural areas, and how it differed for rich and poor

households, and (2) to analyse the differences in the urban-rural welfare gap over time, and what factors contributed towards that gap. This chapter contributes to the literature in two ways. Firstly, the urban-rural welfare gap has not been formally analysed so far in the case of Sri Lanka. Secondly, the chapter adds to the existing literature through the use of a new methodology (the unconditional quantile regression by Firpo et al.) which enabled us to examine the determinants of expenditure and decompose the urban-rural welfare gap across the entire distribution of expenditure. As we observed a rising welfare gap across the distribution, along with the contributions of household characteristics towards this gap, it validated the use of a quantile regression framework – thus, an analysis at the mean would have been inadequate.

Real household expenditure per capita was used as the measure of welfare. In order to identify the determinants of urban and rural household welfare, we used the unconditional quantile regression by Firpo et al. (2009). The determinants of household expenditure suggest that education had a positive and increasing association with per capita expenditure across the distribution. Other determinants positively associated with per capita expenditure include the presence of children, employment in the services sector and receiving remittance income. On the other hand, household size and employment in the agriculture sector were negatively associated with household expenditure.

In the second part of the analysis, the urban-rural welfare gap for 2002 and 2009/10 was estimated using the three-fold Blinder-Oaxaca decomposition. At a given point in time, the welfare gap was larger at higher quantiles suggesting that the gap in expenditure per capita between richer urban and rural households is higher. The factors contributing towards this gap were both, differences in characteristics (such as differences in human capital endowments and regional differences that contribute towards a high welfare gap) and differences in returns to characteristics. In both years, differences in returns to education contributed more to a higher gap at the 25th quantile, but the effect reduces for richer households. Over time, we observed a narrowing of the urban-rural welfare gap. This was due to smaller urban-rural differences in characteristics and the returns to those characteristics in 2009/10. These

characteristics include education and type of employment (wage, self, unemployed/not employed) among others.

1.3 Sectoral choice and sample selectivity

The final part of the thesis estimated the determinants of sector choice, and the returns to employment for individuals in various sectors of the labour market. The data set used for the analysis was the Quarterly Labour Force Survey for 2013. This data set gives detailed information on the income from self-employment which was not available in the previous years of this series. It also gives information on the hours worked which was useful to obtain the earnings per hour (the key variable of interest). Five sectors were identified –non-agriculture public sector employees, non-agriculture private formal employees, non-agriculture private informal employees, non-agriculture self-employed individuals, and agriculture workers (both wage and self-employed). The objectives were three-fold: (1) to address the potential issues of sample selectivity and endogeneity of education, (2) to investigate the determinants of sector choice across various sectors of the labour market, and (3) to establish differences in earnings across the sectors, for men and women. A previous study by Gunatilaka (2008) observed the determinants of sector choice for Sri Lanka in 2006 while dealing with the sample selectivity bias, but was not able to include the self-employed in the analysis and account for the endogeneity bias. This chapter made two contributions to the literature. Firstly, we address two forms of bias – sample selectivity and endogeneity. Secondly, the chapter adds to the existing literature on labour market outcomes in Sri Lanka by including the self-employed and the informal workers in the analysis.

To deal with the issue of sample selectivity, a multinomial logit (MNL) was estimated following Heckman (1979) and Lee (1983). This part of the analysis included individuals who are not in the labour force. The issue of sample selectivity arises due to the non-random nature in which an individual may join a sector of employment. Since we only have information on earnings for those who are already in the labour market, we had to account for those who are not in the labour force – ignoring these individuals may lead to biased estimates of the earnings functions. Further, the MNL

allowed us to identify the determinants of sector choice for individuals who joined the five sectors mentioned above. The results of the MNL estimation suggested that more educated men joined the public and formal private sectors, whereas the least educated men joined the informal private sector. More educated women joined the public sector. Additionally we found that married men are more likely to join the labour force whereas married women are less likely to join. The presence of children in the household did not affect men's decisions to participate in the labour force in most cases, however it had a discouraging effect on women's decisions to participate.

This chapter addressed the potential issue of endogeneity of education. As noted from the findings of the first part of the thesis, it is important to treat education as endogenous. To deal with this bias, a Control Function (CF) approach was used following Garen (1984), Wooldridge (2005), Söderbom et al. (2006) and Kuépié et al. (2009). According to the CF approach, education was regressed on all exogenous variables and the residuals were obtained. The residuals were included in the MNL and earnings functions to account for endogeneity. We then estimated the Mincerian earnings functions for individuals in different sectors of employment, controlling for sample selectivity and endogeneity. The results suggested that endogeneity of education was a bigger issue than sample selectivity in this context – sample selectivity had an effect on men in the formal private sector. Accounting for endogeneity led to higher estimates, in line with the findings previously obtained for the first section of the thesis. The determinants of sector choice and the returns to education varied across sectors validating the importance of disaggregating the labour market into segments; for example, the returns to education varied greatly across the formal and informal private sectors, and this was true for both sexes. Education had a large role to play in facilitating entry into high-paying jobs, such as in the public and formal private sectors. Further, individuals in these sectors received higher returns to education compared to the informal and agricultural sectors.

2 Policy implications, limitations and scope for future research

The study explored the effect of education on earnings and employment determination, and the implications of the recent developments in the Sri Lankan economy. We looked at three areas: (1) the effect of education on earnings and productivity, investigating the issue of endogeneity, (2) the factors contributing towards the urban-rural welfare gap for the years 2002 and 2009/10, and (3) the determinants of sector choice and earnings in a segmented labour market. From the findings, we are able to discuss potential areas of interest for policy makers and scope for further research.

The findings from chapter 2 for the year 2009/10 suggested that education and earnings share a convex relationship, contrary to traditional findings (Psacharopoulos, 1994), however in line with more recent literature (Colclough et al., 2010). Riboud et al. (2007) found Sri Lanka to be an outlier compared to other South Asian countries as more than 70 per cent of the people born in the late 1970s had completed at least primary education. However the country is no longer an outlier when upper secondary education is considered. If the government wishes to focus on equity and efficiency across all levels of education, our results suggest that education policies focussing on post-primary education where the returns are higher might be effective.

The findings from chapter 3 suggested that the urban-rural welfare gap has fallen over time. Additionally, industrialization was favourable towards rural areas in 2009/10. As road infrastructure has improved in the country, it may be a possible explanation for the narrowing of the urban-rural welfare gap. Further research might allow us to identify the improvements in road infrastructure over the years, and how it may have affected the welfare gap. Another issue when estimating the welfare gap for 2002 and 2009/10 was the lack of information on migrants in the data. We however include controls for remittance flows to understand its association with household expenditure. Being able to separate the migrants and understand the movements over time is another area for future research.

The findings from chapter 4 suggested that individuals with higher levels of education are more likely to join high-paying jobs such as public or formal private sector employment, and less likely to join the informal and agriculture sectors. However, over 60 per cent of the total labour force is employed in the informal sector. Given that education is an important factor affecting individual's sector choices, understanding why certain people get less education than others is a potential area for further research. A limitation with the analysis in chapter 4 was that it was carried out at the mean, rather than exploring the entire earnings distribution. This was due to the methodology used in the chapter that tried to deal with both, the endogeneity and sample selectivity biases. Additionally, the analysis did not include controls for occupation because it does not seem like the choice of occupation and sector occur sequentially. For future research on this topic of labour market outcomes in Sri Lanka, we recommend an extension of the study on sector choice to include controls for occupation and examination of the entire earnings distribution.

Despite these limitations, this study adds to the existing literature on labour market outcomes. Firstly, we identified the importance of treating education as endogenous. We also observed the rising returns to education at upper secondary and tertiary levels. However, the household production function suggested that higher education is positively associated with non-agricultural output, but not with agricultural output. Secondly, we observed a falling urban-rural welfare gap over time in Sri Lanka due to smaller urban-rural differences in characteristics and returns to characteristics in 2009/10, compared to 2002. Finally, we noted that the determinants of sector choice and the returns to education vary across men and women. While women earn more than men, on average, in the public sector, this is not the case in other sectors of employment. Further, we observed that married women are less likely to join the labour force while married men are more likely to join.