CREDIT Research Paper

No. 18/09

Sectoral choice and selectivity

by

Nirodha Bandara, Simon Appleton and Trudy Owens

Abstract

This paper attempts to examine the labour force participation decisions and earnings across employment sectors and how it varies by gender in Sri Lanka. The labour market is disaggregated into 5 sectors – public, formal private, informal private, self-employed and agriculture. Using the Labour Force Survey 2013, this paper adds to existing literature in two ways. Firstly, the paper deals with two forms of potential biases which have not been simultaneously explored for the case of Sri Lanka – sample selectivity and endogeneity of education in earnings. Secondly, it adds to the literature by including the self-employed in the analysis. The determinants of sector choice are analysed using a multinomial logit. The findings of this paper suggest that individuals with the highest levels of education get into 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. The differences across sectors confirm the importance of disaggregating the sectors of employment to examine choices of labour force participation.

JEL Classification: C3, J24, O12

Keywords: sample selectivity, endogeneity, multinomial logit, control function, Sri Lanka

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

CREDIT Research Paper

No. 18/09

Sectoral choice and selectivity

by

Nirodha Bandara, Simon Appleton and Trudy Owens

The Authors

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

Acknowledgements

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

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

1. Introduction

This paper examines the 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 paper deals with two forms of possible biases which have not been examined simultaneously in the context of Sri Lanka in previous literature[1] – (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.

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 countries traditionally focuses 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 paper stems from this area of the literature.

The paper 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. 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 paper is threefold. First, to analyse the determinants of sector choice and earnings across sectors we address the issues of sample selectivity and

1 This was revealed from a literature search in EconLit

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 paper 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 paper 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 focuses on econometric issues raised by the various studies that are subsequently used to guide the empirics of this paper.

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, 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 paper 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.

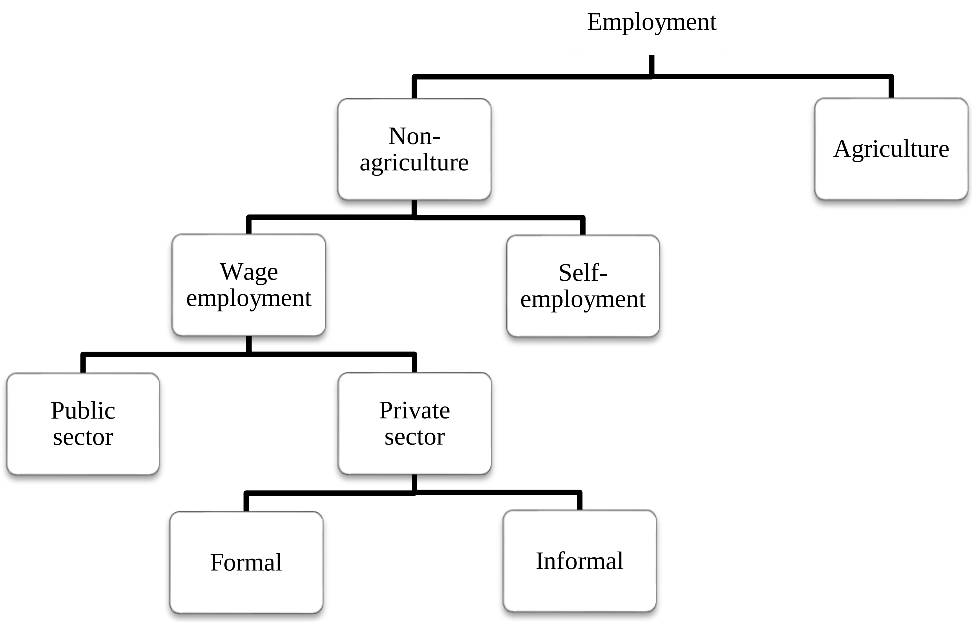
Before exploring the summary statistics, it is important to clearly differentiate between formal and informal private sector employment. The definition of informal employment may vary from study to study, often due to data constraints. Several papers broadly use the definition adopted at the 15th International 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.

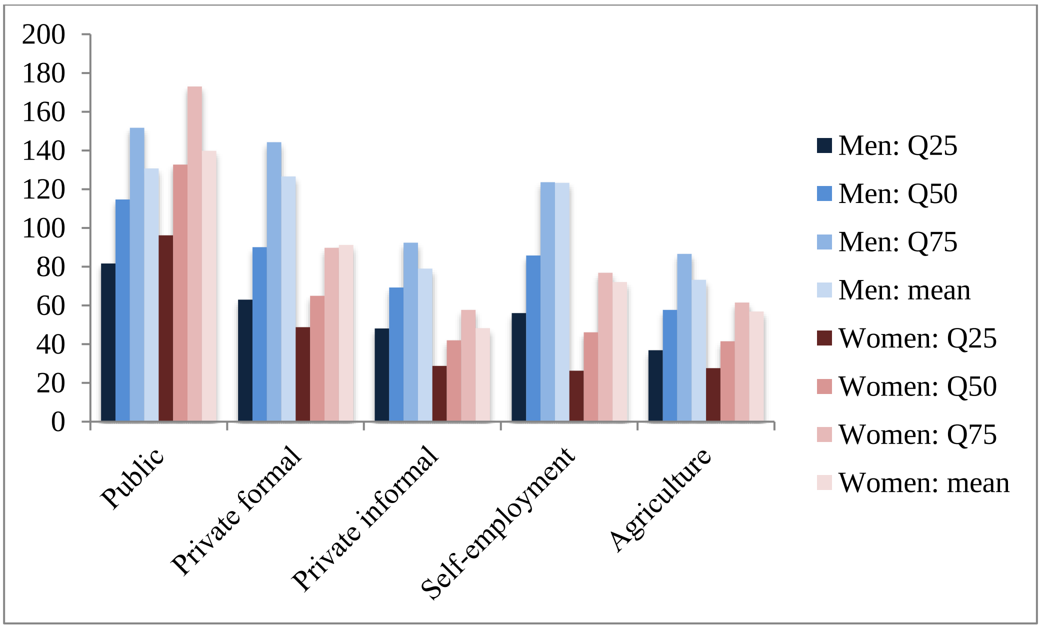
The above characteristics of the informal sector strengthen the reasoning for separating the informal and formal sectors. As the characteristics of and returns to wage employment may vary from self-employment, this study distinguishes between the two categories. The divisions of employment used in this 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[2]. The earnings per hour are calculated as reported weekly earnings divided by the reported actual hours worked during the week[3]. 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

2 The descriptive statistics for the full sample are presented in the Appendix Table A1.
3 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
<u>Industry (base: services)</u>												
Manufacturing	0.07	0.46	0.44	0.27			0.03	0.65	0.49	0.47		
<u>Occupation (base: other)</u>												
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:

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

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 paper. 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 paper 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 $\varphi(h_{ij})$ 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[4]:

(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,

4 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.

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)[5].

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 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[6]. 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

5 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.

6 The results are presented in the Appendix Table A3

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 paper also deals with the possible endogeneity of education - that is, an individual's level of education could be correlated with the error term (ζ_{ij}) in equation (4) because of unobservable individual heterogeneity. Failing to account for the possible endogeneity of education 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 paper 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[7].

7 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:

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

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[8]. These include the spouse/mother/father's level of education 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

8 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).

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[9]. 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

9 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).

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.

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	5.97×10^{-5}***	3.40	1.35×10^{-4}***	10.30	1.31×10^{-4}***	21.22	1.43×10^{-4}***	22.46	1.12×10^{-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	5.11×10^{-5}**	-2.55	1.14×10^{-4}***	10.42	8.67×10^{-5}***	10.57	4.64×10^{-5}***	5.37	6.39×10^{-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 $\leq$ 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[10], 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[11].

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

11 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 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 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 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 paper 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: Earnings 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	1.21×10^{-5}	0.90	1.42×10^{-6}	0.20	1.44×10^{-5}***	3.53	5.00×10^{-6}	0.71	7.44×10^{-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	-1.16×10^{-4}	-0.12	-0.002	-1.04	-0.001	-0.47
Age^3	7.57×10^{-6}	0.32	2.71×10^{-6}	0.21	5.19×10^{-8}	0.01	8.38×10^{-6}	0.78	5.45×10^{-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[12]. 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.

12 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 paper 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 paper 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 paper 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. 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 paper 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.

References

Adamchik, V. and Bedi, A., 2000, "Wage Differentials between the Public and Private Sectors: Evidence from an Economy in Transition", *Labour Economics*, 7, 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", *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 Bace Publishers, Evidence and Policies, Fortworth

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

Jaan, B., 2008, “The Blinder-Oaxaca decomposition for linear regression models”, *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

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)	0.05	0.54	0.46	0.33		
Manufacturing						
Occupation						
(base: other)	0.53	0.31	0.08	0.21	0.01	
High skilled	0.24	0.19	0.20	0.22		
white collar	0.03	0.21	0.30	0.34	0.57	
Low skilled	0.16	0.29	0.42	0.22	0.42	
white collar						
High skilled						
blue collar						
Low skilled						
blue collar						
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	$1.28\text{x}10^{-4}$***	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 ≤ 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	9.49×10^{-6}	0.87	5.53×10^{-6}	0.87	1.29×10^{-5}***	3.70	7.56×10^{-6}	1.31	6.20×10^{-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	1.21×10^{-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	6.28×10^{-5}	0.07	-0.002***	-4.08	-0.001	-1.24	-0.001	-1.62
Age^3	9.99×10^{-6}	0.75	-5.57×10^{-6}	-0.77	1.42×10^{-5}***	3.50	4.21×10^{-6}	0.60	7.59×10^{-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	-9.43×10^{-6}	-0.01	-1.02×10^{-4}	-0.10	-0.002	-0.91	-6.93×10^{-4}	-0.37
Age^3	9.44×10^{-6}	0.39	-1.33×10^{-6}	-0.10	-2.48×10^{-7}	-0.04	6.99×10^{-6}	0.66	4.368×10^{-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	4.17×10^{-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 (781)	112.82 (1316)	86.76 (1739)	138.14 (853)	102.06 (1019)
> 30	162.88 (3272)	147.60 (1723)	92.74 (2916)	154.17 (5202)	97.05 (6561)
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 (208)	91.89 (285)	89.03 (1096)	133.69 (1243)	100.47 (2112)
Upper secondary	142.94 (2810)	132.05 (2401)	94.78 (2765)	166.16 (3875)	106.68 (2576)
Tertiary	203.23 (944)	250.86 (202)	119.58 (41)	263.53 (118)	122.51 (41)

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. 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 $\beta*$ is a non-discriminatory coefficient vector[13].

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

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.

[13] There are several ways in which $\beta*$ 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 $\beta*$; relative weights are given to the coefficients of each group.

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[14].

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.

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

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.

<u>Table A11: Blinder-Oaxaca decomposition</u>

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)	$3.97\text{X}10^{-4}$	0.51	0.011***	2.86	$-4.64\text{x}10^{-4}$	-0.16	0.014***	3.45	0.006	1.58
Ethnicity	$-5.54\text{X}10^{-5}$	-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)