Estimation of wage equations for New Zealand

Joseph Mercante and Penny Mok

New Zealand Treasury Working Paper 14/09

April 2014

THE TREASURY
Kaitohutohu Kaupapa Rawa

New Zealand Government

NZ TREASURY WORKING PAPER 14/09

Estimation of wage equations for New Zealand

MONTH/YEAR

April 2014

AUTHORS

Joseph Mercante
New Zealand Treasury
1 The Terrace
Wellington
New Zealand
*Joseph Mercante has since returned to the Australian Treasury.

Email joseph.mercante@treasury.gov.au
Telephone +61 2 6263 3370
Fax +61 2 6263 4322

Penny Mok
New Zealand Treasury
1 The Terrace
Wellington
New Zealand

Email penny.mok@treasury.govt.nz
Telephone +64 4 917 6274
Fax +64 4 473 0982

ISBN (ONLINE)

978-0-478-42163-7

URL

Treasury website at April 2014:
http://www.treasury.govt.nz/publications/research-policy/wp/2014/14-09
Persistent URL: http://purl.oclc.org/nzt/p-1646

ACKNOWLEDGEMENTS

Access to the data used in this paper was provided by Statistics New Zealand in accordance with security and confidentiality provisions of the Statistics Act 1975. We would like to thank Guyonne Kalb, John Creedy, Gerald Minnee, Nick Carroll and referees for their helpful comments and assistance. This paper was undertaken while Joseph Mercante was on secondment to the New Zealand Treasury.

NZ TREASURY

New Zealand Treasury
PO Box 3724
Wellington 6008
NEW ZEALAND

Email information@treasury.govt.nz
Telephone 64-4-472 2733
Website www.treasury.govt.nz

Abstract

This paper estimates wage equations for New Zealand based on pooled data from the Household Economic Survey (HES) from 2006/07 to 2010/11. Equations are estimated separately for couple men and women, single men and women and sole parents. The results are compared to previous New Zealand estimates using the HES from 1991 to 2001. We estimate wage equations and account for possible sample selection bias. Our estimates of the wage equations are largely comparable to the earlier estimates. We extend the equations to predict wage rates for workers and non-workers by drawing from the estimated wage distribution. We find that sample selection is significant for married men, single women and sole parents but not significant for married women and single men. Overall, we find that wage rates are positively related to age, education and experience but the wage rates are generally lower for non-Europeans and for people living outside Auckland.

JEL CLASSIFICATION J21
J31

KEYWORDS wage rate; labour supply; selection bias; wage prediction

Executive Summary

The purpose of this paper is to estimate wage equations for New Zealand based on the Household Economic Survey (HES) from 2006/07 to 2010/11. This paper updates wage estimates done a decade earlier.

The main motivation for the paper is to capture changes in the New Zealand economy over the decade since the equations were last estimated. Over the past decade, there has been an increase in education of the workforce and a reduction in the gender earnings disparities over time which has created a positive environment for women to participate in the labour market.

The estimated equations are ultimately used in the behavioural microsimulation model (TAXMOD-B) to predict labour supply responses to policy changes. The equations are used to impute wage rates for those who are currently not working.

The wage equations are estimated separately for couple men and women, single men and women and sole parents and take account of possible sample selection bias. We extend our analyses in predicting the wage rates for workers and non-workers by drawing from the estimated wage distribution, which incorporates variation due to unobserved heterogeneity.

In estimating the wage equations we take account of education, ethnicity, geography and industry and occupation. We also take account of the trend in wages over time and the business cycle by including controls for the unemployment rate and a time trend variable. In controlling for sample selection we take account of other characteristics such as the presence and ages of children, the presence of unearned income and for couples, whether the partner works and other selected characteristics of the partner.

Our results generally confirm earlier research. We find that sample selection is significant for married men, single women and sole parents while it is not significant for married women and single men. We also find that the wage rates are positively related to age, education and experience while it is lower for non-Europeans and for people living outside Auckland. Generally, people at higher education levels experience a steeper wage increase and reach the maximum rate at an older age. The turning points where wage rates starts to decrease with age occur later in the mid-forties or early fifties compared to estimates made a decade earlier. This could be explained by increasing education attainment over time in New Zealand where the earnings of the better-educated workers rise more quickly because they are investing in job training and the rise will be for a longer time than their less-educated counterparts.

There are considerable differences in wage rates between occupations, industries, regions and ethnicity for all demographic groups. People working in managerial and professional positions experience higher wages than people working in elementary occupations. People working in the Finance and Public sectors are paid higher wages than the Services sector and people living in Auckland are the highest wage earners. Individuals from European descent are paid more than individuals from other ethnicities, though this is not significant for sole parents.

We use the wage equations to predict conditional hourly wage rates for individuals in our sample. In predicting wages of workers and non-workers, we find that the mean wage has a narrower distribution than the distribution of observed wages. We then predict wages with a random disturbance term to introduce a wider dispersion in the wage distribution. This is particularly useful in labour supply simulation models which typically require wage rates to be imputed for non-workers. However, the greater dispersion of the predicted wage distribution when including unobserved heterogeneity can result in wages below the statutory minimum wage. We find this to be the case particularly for sole parents.

For single men we found that the sample selection model at times produced counter-intuitive wage predictions. The problem arises because of a negative sample selection term.

The present paper is the first in a suite of papers that will emerge from NZ Treasury's behavioural microsimulation modelling project. The second paper estimates the labour supply equations used to predict labour supply responses to changes in financial incentives. The third paper analyses labour supply responses to selected tax and benefit policies in New Zealand. Together, the papers will offer substantive evidence on labour supply responses to tax and benefit policy changes in New Zealand.

Table of Contents

List of Tables

List of Figures

Estimation of wage equations for New Zealand

1 Introduction

The New Zealand labour market has experienced dramatic changes over the past decades. Some of the changes are evidenced by increasing educational attainment of the workforce and the reduction in gender earnings disparities over time which has created a positive environment for women to participate in the labour market. The proportion of New Zealanders aged 25 to 34 years with a bachelors or higher qualification has expanded twofold over the last decade, from 16% in 2001 to 30% in 2011 (Ministry of Education, 2012). Women are participating in tertiary education at a much higher rate than men. As a result, the gap in the labour force participation rates between men and women has decreased for those with tertiary qualifications and thus reduced the earnings disparities for women with tertiary qualifications. These trends are likely to have an effect on earnings, especially of women.

The aim of this paper is to estimate the wage equations used to impute wage rates in our behavioural microsimulation model (TAXMOD-B) in the light of the recent changes in the labour market. Earlier versions of the wage equations were estimated based on the 1991/92 to 2000/01 HES (see Kalb and Scutella, 2003). A behavioural microsimulation model predicts labour supply responses to policy changes and is useful when assessing policy changes because many policy changes (ie. tax changes) are designed with the aim of altering the behaviour of individuals towards greater participation in the labour market. The model requires a large cross-sectional dataset which contains the characteristics of individuals and households, their labour supply and earnings. In this dataset, information about the potential wage of non-workers is missing, and therefore the need to impute wages for these people. For the behavioural microsimulation model to produce reliable results, it is helpful that the estimates are based on the most up-to-date data.

This paper estimates wage equations separately for five demographic groups – married men, married women, single men, single women and sole parents in New Zealand using pooled information from the 2006/07 to 2010/11 Household Economic Survey (HES). The availability of five data sets covering a period of six years allows us to capture the more recent changes in the labour market that affect wages. We estimate wage equations using two sample selection models: the two-step Heckman procedure and a maximum likelihood procedure which simultaneously allows for correlation between the unobserved components of the wage and employment equations. Our discussion focuses primarily on the latter which makes the most efficient use of the available data. We extend our analyses in predicting wage rates for workers and non-workers by drawing from the

estimated wage distribution. This allows the incorporation of variation in the predicted wage due to the unobserved heterogeneity in the data. The results of predicted wages with and without the additional variability are compared.

The present paper is the first in a suite of papers that will emerge from NZ Treasury's behavioural microsimulation modelling project. The second paper estimates the labour supply equations used to predict labour supply responses to changes in financial incentives. The third paper analyses labour supply responses to selected tax and benefit policies in New Zealand. Together, the papers will offer substantive evidence on labour supply responses to tax and benefit policy changes in New Zealand.

The remainder of the paper is organised as follows. The next section provides a brief overview of the changes over the past years that have impacted on wages. Section 3 describes the econometric methodology to estimate wage equations. Section 4 briefly describes the data. The estimates of the wage equations are reported in Section 5. Section 6 discusses the prediction of wage rates for certain demographic groups. The last section concludes.

2 Background

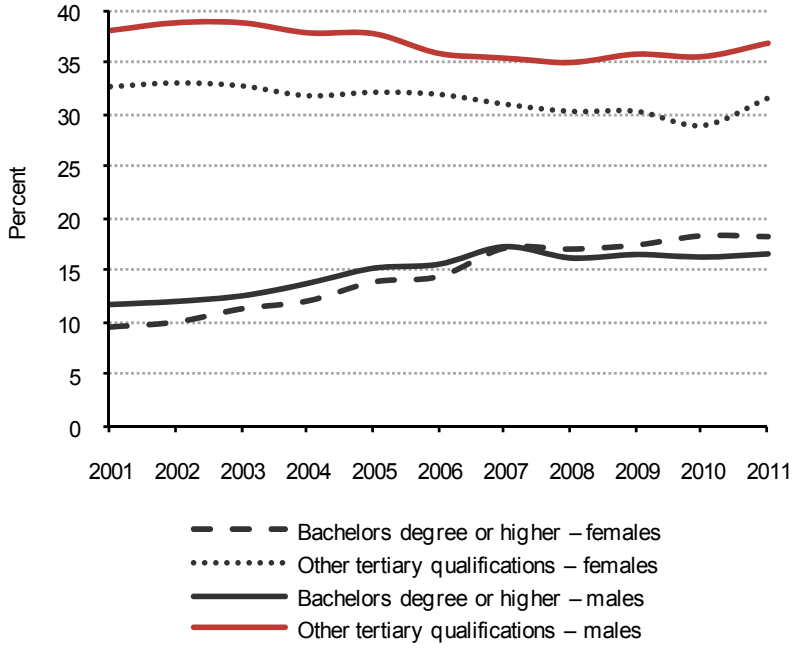
This section provides a brief and non-exhaustive overview of the economic changes experienced in the country over the past decades that have had a significant impact on wages. The impact has predominantly arisen through an increased educational attainment of the workforce and rising workforce participation of women. Hyslop and Mare (2009) found that educational attainment among all New Zealand residents increased markedly between 1986 and 2001. They found that the fraction of workers with university degrees doubled from 7.5 to 15% while the fraction with no qualifications fell from one-third to less than 20% using the census data. Similar patterns are documented by the Ministry of Education. In addition, the proportion of New Zealanders aged 25 to 34 years with a bachelors or higher qualification has expanded twofold over the last decade, from 16% in 2001 to 30% in 2011 (Ministry of Education, 2012).

Historically, men have had a higher level of tertiary qualification attainment than women in New Zealand. However, the gap in attainment between the genders is closing. One of the contributing factors to the closing of the gap in the attainment of tertiary qualifications by men and women is that women are participating in tertiary education at a much higher rate than men. Figure 1 shows that the proportion of the population holding a bachelors or higher qualification is higher for women while men are more likely to hold tertiary certificates and diplomas. In 2001, 51% of men had a tertiary-level qualification[1], compared with 43% of women. In 2011, 54% of men and 50% of women had attained a tertiary qualification. This represents a closing of the gap in attainment from 8 percentage points in 2001 to 4 percentage points in 2011 (Ministry of Education, 2012). In addition, the chances of completing a qualification are much greater for women than men. The seven-year completion rate for women who started a tertiary qualification in 1998 was 48%, compared with 37% for men (Smart, 2006).

Although women's participation rate in the labour market is lower than men's, the gap between them is decreasing for people with higher-level qualifications. There is a substantial body of evidence that shows that those with higher levels of education are more likely to participate in the labour market, have greater access to further training and receive higher earnings on average. In other words, tertiary qualifications improve the access of women to the labour market. Figure 2 shows that in 2005, men with a bachelors or higher qualification had a labour force participation rate of 87%. Women with the same level of qualification had a labour force participation rate of 80%, a gap of 7 percentage points. This compares with a gap of 15 percentage points for the same groups in 1991. In contrast, the participation rates for men and women with no qualifications are much lower at 60% and 39% respectively in 2005 (Smart, 2006).

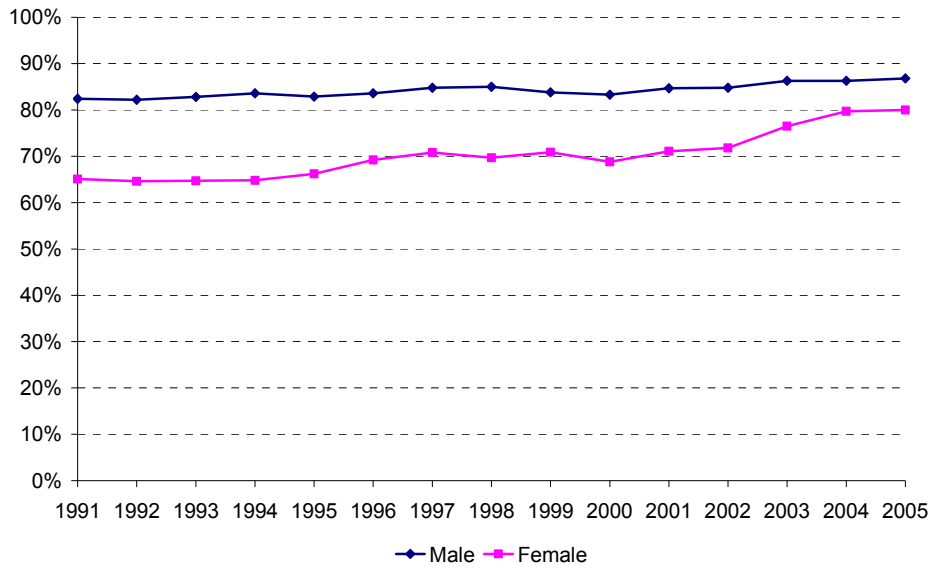
[1] Tertiary qualifications are divided into 2 categories: 'bachelors or higher', and 'other tertiary'. 'Bachelors or higher' includes postgraduate degrees, certificates or diplomas. 'Other tertiary' includes university certificates or diplomas, teaching certificates or diplomas, nursing certificates or diplomas, NZ certificates or diplomas, technician certificates, local polytechnic certificates or diploma, trade certificates or advanced trade certificates.

Figure 1 – Population aged 15 years and over (June quarter) with a tertiary qualification by gender



Source: Ministry of Education (2012), Statistics New Zealand- Household Labour Force Survey.

Figure 2 – Labour force participation rate of the population aged 15 and over with a bachelors or higher qualification by gender 1991-2005



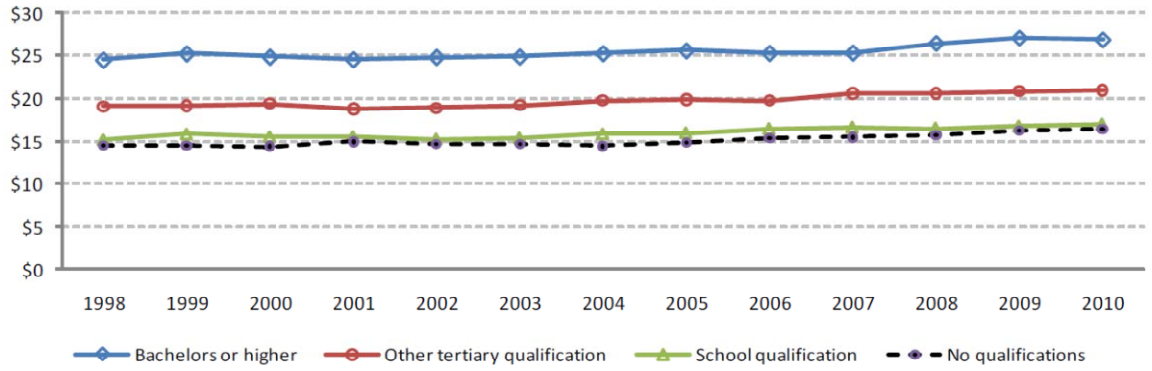
Source: Smart (2006), Statistics New Zealand- Household Labour Force Survey.

Data from Statistics New Zealand's Income Survey show a strong association between the attainment of tertiary qualifications and higher incomes. Maani (1999) analysed New Zealand Census data between 1981 and 1996 and found that there was a significant premium on attaining tertiary qualifications. The premium on the qualifications was higher in 1996 than it was in 1981, despite an increase in the proportion of people in the New Zealand labour force holding a tertiary qualification over that time.

In 2000 to 2005 when labour demand was strong, the earnings of people with a bachelors or higher qualification were 70% higher than for people without a qualification. In 2010, the earnings difference between the same groups of people was 63% but had increased to 65% in 2011, displaying a slight recovery in the labour market after the economic downturn (Ministry of Education, 2012).

Figure 3 shows that in 2010 the median hourly earnings for those with tertiary degrees was $26.85, approximately 1.6 times higher than those with school qualifications. On average, earnings for those with tertiary education are 18% higher compared to those with upper secondary and post-secondary non-tertiary education (Ministry of Education, 2011).

Figure 3 – Real median hourly earnings by highest qualification for people aged 15 and over (2010 NZ dollars)



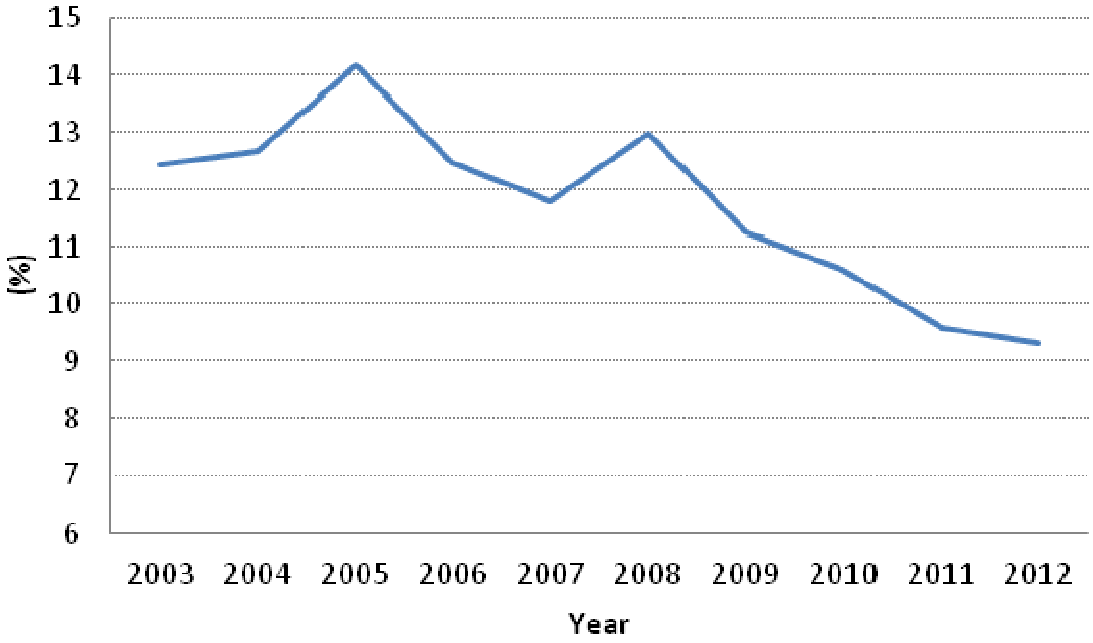
Source: Ministry of Education (2011), Statistics New Zealand- New Zealand Income Survey.

Disaggregation of earnings by gender shows that the percentage gain for men having bachelor degrees or higher qualifications compared with those with no qualifications decreased from 197% in 1997 to 149% in 2007. Over the same period, the percentage gain for women in the same education categories rose from 132% to 145%.

Gender disparities in earnings exist in New Zealand, with men generally earning more than women who hold the same qualification. However, the disparities have narrowed for the past few years. The overall gender pay gap estimated using median hourly earnings decreased from 12% in 2003 to 9.3% in 2012 (Ministry of Women's Affairs). Refer to Figure 4.

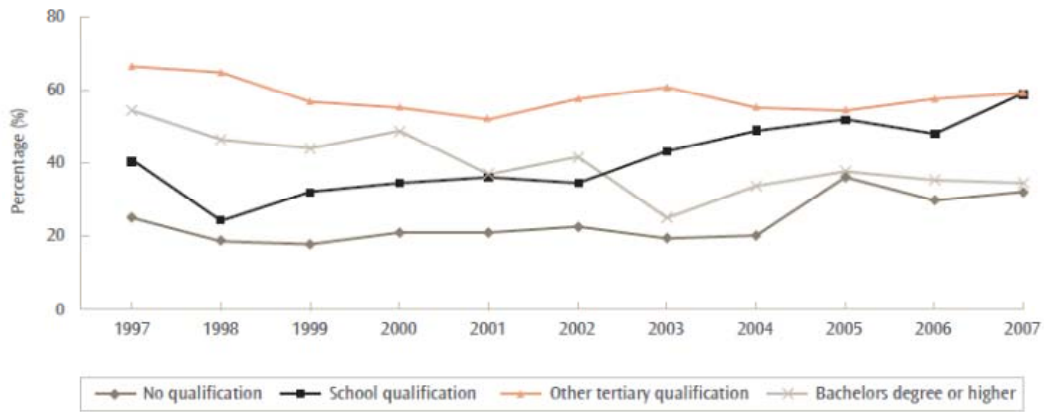
The gender gap in earnings within the same qualification level are attributable partly to the choice of fields of study, industry of employment and time spent in the labour force. Dixon (2000) found that gender differences in educational attainment contributed to the pay gap but the contribution was smaller and declined in importance as the gender educational differences narrowed. The premium in median weekly income experienced by men over women decreased from 67% in 1997 to 42% in 2007. While the gender earnings gap decreased slightly for those with other tertiary qualifications, it increased for those with school qualifications or no qualification (Ministry of Education, 2008). Refer to Figure 5. More women attaining tertiary education and the reduction in the gender earnings disparities over time for those with tertiary qualifications have created a positive environment for women to participate in the labour market.

Figure 4 **–** Gender pay gap (%) 2003-2012



Source: Ministry of Women's Affair (2013), Statistics New Zealand- New Zealand Income Survey.

Figure 5 **–** Earnings premium of men over women by highest qualification median weekly income (1997-2007)



Source: Ministry of Education (2008), Statistics New Zealand .

3 Statistical Model

The estimation of wage equations for the five population groups follows closely the method described in Kalb and Scutella (2003), and is summarised below.

Estimating the wage model

The wage rate is only observed for workers. Estimating wage equations by standard linear regression on the sample of workers is likely to result in sample selection bias. In this analysis we treat all non-workers the same and do not distinguish between the unemployed and not in the labour force.

A common approach to addressing sample selection bias is to estimate wage equations as part of a system of two correlated equations. One equation, the selection equation, determines the selection into employment. Another equation, the wage equation, determines the hourly wage rate conditional on the person being employed. Unless otherwise qualified, throughout this paper the terms wage and wage rate refer to the hourly wage rate.

The participation equation is defined:

$$E_i = \begin{cases} 1 \text{ if } E_i^* > 0 \\ 0 \text{ if } E_i^* \leq 0 \end{cases} \tag{1}$$

where E_i is employment and individual *i* is observed to be employed ($E_i = 1$) or not-working ($E_i = 0$). We assume that our observation of E_i is being determined by some underlying unobserved process called employability E_i^* that measures the tendency to participate in employment and is defined by:

$$E_i^* = z_i^{'}\gamma + u_i \tag{2}$$

where E_i^* may be interpreted as relating to the probability that a wage offer to a person is greater than the reservation wage (a worker) or a wage offer to a person is less than or equal to the reservation wage (a non-worker). A person is observed to be employed if $E_i^* > 0$ and not-employed otherwise. The vector γ is a vector of characteristics that influence the probability of employment.

If we assume that $u_i \sim N(0,1)$, then the probability of working ($E_i^* > 0$) has a probability $\Phi\,(z_i^{'}\gamma)$ and the probability of not working ($E_i^* \leq 0$) has a probability $1\text{-}\Phi(z_i^{'}\gamma)$, where Φ is the standard normal cumulative distribution function.[2]

[2] For the probability of working

$$\Pr(E_i^* > 0) = \Pr(z_i'\,\gamma + u_i > 0) = \Pr(\frac{u_i}{\sigma_u} > -\frac{z_i'\,\gamma}{\sigma_u}) = 1 - \Phi\,(-z_i'\,\gamma) = \Phi\,(z_i'\,\gamma) \text{ since } \sigma_u = 1$$

and synonymously for the probability of not working.

The equation for the wage rate is defined as:

$$W_i = \begin{cases} \text{W if } E_i^* > 0 \\ \text{not observed if } E_i^* \leq 0 \end{cases} \tag{3}$$

where W_i is only observed if a person is employed, and is unobserved for non-workers. Conditional on working, the wage equation is represented by (4) where w_i is the log wage rate, and β is a vector of human capital characteristics that influence the individual's market wage rate.

$$\mathrm{w}_i = x_i^{'}\beta + \varepsilon_i \tag{4}$$

The conditional mean of (4) is shown to be:

$$\begin{aligned} &\mathrm{E}(\mathrm{w}_i \mid E_i^* > 0) = E(x_i^{'}\beta + \varepsilon_i \mid E_i^* > 0) \\ &= E(x_i^{'}\beta + \varepsilon_i \mid z_i^{'}\gamma + u_i > 0) \\ &= x_i^{'}\beta + E(\varepsilon_i \mid u_i > -z_i^{'}\gamma) \end{aligned} \tag{5}$$

If the errors of equations (2) and (4), u_i and ε_i, are independent then $E(\varepsilon_i \mid u_i > -z_i^{'}\gamma)$ =0 and $\mathrm{E}(\mathrm{w}_i \mid E_i^* > 0) = x_i^{'}\beta$ and therefore an OLS regression of *w* on *x* over the sample of workers would be consistent. However, if ε_i and u_i are correlated then OLS on the selected sample would produce biased estimates and an alternative method would be needed to account for selection bias (Cameron and Trivedi, 2005).

The bivariate sample selection model can be used to consistently estimate the coefficients of the wage model.[3] We proceed by assuming that the errors u_i and ε_i are jointly normally distributed:

$$\begin{bmatrix} u_i \\ \varepsilon_i \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_\varepsilon^2 \end{bmatrix}\right] = N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & \sigma_\varepsilon^2 \end{bmatrix}\right] \tag{6}$$

where the correlation between the error terms is $\rho = \dfrac{\sigma_{12}}{\sigma_\varepsilon \sigma_u} = \dfrac{\sigma_{12}}{\sigma_\varepsilon}$ since $\sigma_u^2 = 1$.

The model can be estimated by maximum likelihood. For workers (ie. $E_i^* > 0$) we observe w_i with a probability of $f(w_i \mid E_i{*} > 0).\Pr(E_i{*} > 0)$ where the first term is the conditional distribution of w_i given it is observed and the second term is the probability that w_i is observed. The function $f()$ is the univariate normal density function *φ(·)*. Also the probability function $\Pr(E_i{*} > 0)$ is the standard normal cumulative distribution function $\Phi(\cdot)$. The probability of being a non-worker is $\Pr(E_i{*} \leq 0)$ where $\Pr(E_i{*} \leq 0)$ is one minus the standard normal cumulative distribution function $1 - \Phi(\cdot)$. The likelihood function is then given by (7).

[3] The bivariate sample selection model is known by various names such as the "Tobit 2 model" or simply the "selection model with a probit selection equation". See Cameron and Trivedi (2005) for a discussion.

$$L = \prod_{i}^{n} \left\{ \left[\Pr(E_i^* \le 0) \right]^{1-E_i} \cdot \left[f(w_i \mid E_i^* > 0) \cdot \Pr(E_i^* > 0) \right]^{E_i} \right\} \tag{7}$$

where E_i is as defined in (1). Estimation of this equation relies on strong distributional assumptions for the error terms.[4] An alternative less efficient estimator is the Heckman two-step sample selection model originally proposed by Heckman (1979). That estimator makes weaker distributional assumptions than the maximum likelihood estimator. We estimated the wage equations using both methods (maximum likelihood and Heckman's two-step), however in line with the reasoning in Kalb and Scutella (2003), we report on the maximum likelihood results given that the simultaneous wage and selection model estimated by maximum likelihood makes more efficient use of the available data.

Equation (7) involves terms in $x_i' \beta$ and $z_i' \gamma$. It is generally recommended that for identification purposes, at least one regressor of the explanatory variables in the participation equation z_i' is excluded from the explanatory variables of the outcome (that is wage) equation x_i' (Cameron and Trivedi, 2005).

Predicting wages

The conditional mean wage in (5) can be simplified taking note of the relationship between ε_i and u_i as $\varepsilon_i = \sigma_{12} u_i + \xi_i$ with u_i and ξ_i independent and so $\operatorname{cov}(u_i, \xi_i) = 0$. We can then write (5) as[5]:

$$\begin{aligned} \mathrm{E}(\mathrm{w}_i \mid E_i^* > 0) &= x_i' \beta + E(\varepsilon_i \mid u_i > -z_i' \gamma) \\ &= x_i' \beta + E(\sigma_{12} u_i + \xi_i \mid u_i > -z_i' \gamma) \\ &= x_i' \beta + \sigma_{12} E(u_i \mid u_i > -z_i' \gamma) \end{aligned} \tag{5a}$$

The last expression is an error term from the truncated normal distribution.[6] In general, for any standard normal distribution $\mathrm{z} \sim \mathrm{N}(0,1)$, the truncated mean $E(z \mid z > -c) = \dfrac{\phi(c)}{\Phi(c)}$ [7] then substituting into (5a) becomes:

$$\begin{aligned} \mathrm{E}(\mathrm{w}_i \mid E_i^* > 0) &= x_i' \beta + \sigma_{12} E(u_i \mid u_i > -z_i' \gamma) \\ &= x_i' \beta + \sigma_{12} \frac{\phi(z_i' \gamma)}{\Phi(z_i' \gamma)} \\ &= x_i' \beta + \sigma_{12} \lambda(z_i' \gamma) \\ &= x_i' \beta + \rho \sigma_{\varepsilon} \lambda(z_i' \gamma) \end{aligned} \tag{8a}$$

---

4 For further discussion of the estimation of the likelihood function and the distributional properties see Cameron and Trivedi (2005) and Amemiya (1985).

5 This is shown in Cameron and Trivedi (2005).

6 Note that in the expression $E(u_i \mid u_i > -z_i' \gamma)$ we divide by σ_u to get $E(u_i \mid \frac{u_i}{\sigma_u} > \frac{-z_i' \gamma}{\sigma_u})$ where $\sigma_u = 1$.

7 Truncated means with truncation from below and above follow directly from Theorem 22.2 in Greene (2003).

where $\lambda(z_i' \gamma)$ is called the inverse Mill's ratio. This shows that the expected wage conditional on working is an amount more than $x_i'\beta$ if $\rho\sigma_\varepsilon$ is positive. Similarly, for non-workers, using $E(z \mid z \le -c) = \frac{-\phi(c)}{1-\Phi(c)}$ and substituting into (5a) we obtain:

$$\begin{aligned} \mathrm{E}(\mathrm{w}_i \mid E_i^* \le 0) &= x_i'\beta + \sigma_{12} E(u_i \mid u_i \le -z_i'\gamma) \\ &= x_i'\beta - \sigma_{12} \frac{\phi(z_i'\gamma)}{1-\Phi(z_i'\gamma)} \\ &= x_i'\beta - \rho\sigma_\varepsilon \frac{\phi(z_i'\gamma)}{1-\Phi(z_i'\gamma)} \end{aligned} \tag{8b}$$

The expressions of equations (8a) and (8b) are conditional mean wages. They are conditional because we are conditioning on the wage being either observed (workers) or unobserved (non-workers). In contrast, the unconditional mean wage is when we do not condition on whether the person is observed working or not working and from (4) is given by:

$$\mathrm{E}(\mathrm{w}_i) = E(x_i'\beta + \varepsilon_i) = x_i'\beta \tag{9}$$

Having estimates $\hat{\beta}, \hat{\rho}, \hat{\sigma}_\varepsilon, \hat{\gamma}$ and $\hat{\lambda}_i$ the conditional mean wage predictions for workers and non-workers are obtained from (8a) and (8b) and are respectively:

$$E(\hat{w}_i \mid_{E_i=1}) = x_i'\hat{\beta} + \hat{\rho}\hat{\sigma}_\varepsilon \hat{\lambda}_i \tag{10a}$$

$$E(\hat{w}_i \mid_{E_i=0}) = x_i'\hat{\beta} - \hat{\rho}\hat{\sigma}_\varepsilon \left[\frac{\varphi(z_i'\hat{\gamma})}{1-\Phi(z_i'\hat{\gamma})} \right] \tag{10b}$$

Given a positive estimate of $\hat{\rho}\hat{\sigma}_\varepsilon$, for two persons with the same characteristics, the predicted mean wage of a working individual is higher than the non-working individual. However, as discussed by Ermisch and Wright (1994), a negative value of lambda is not implausible and is not necessarily an indication of a misspecification.

Finally, the unconditional mean wage prediction from (9) is given by:

$$E(\hat{w}_i) = x_i'\hat{\beta} \tag{11}$$

Predicted wage with random disturbance term

We expect the conditional mean wage distributions predicted from (10a) and (10b) to be fairly narrow as they assign the same mean wage to persons with similar characteristics. To introduce greater heterogeneity in the wage prediction, we draw from the estimated distribution of error terms which represents unobserved heterogeneity.

The predicted wage for workers and non-workers can be computed as follows:

$$\mathrm{w}_i \mid_{E_i=1} = E(w \mid_{E_i=1}) + \upsilon_i = x_i'\beta + \rho\sigma_\varepsilon \lambda_i + \upsilon_i \tag{12a}$$

$$\mathrm{w}_i \mid_{E_i=0} = E(w \mid_{E_i=0}) + \upsilon_i = x_i'\beta - \rho\sigma_\varepsilon \left[\frac{\varphi(z_i'\gamma)}{1-\Phi(z_i'\gamma)} \right] + \upsilon_i' \tag{12b}$$

The variance of υ_i and $\upsilon_i^{'}$ are heteroskedastic as described in Kalb and Scutella (2003) and Greene (2003) which include a discussion of calculating this variance from estimates of $\hat{\rho}, \hat{\sigma}_\varepsilon$ *and* $\hat{\lambda}_i$.[8]

The error terms υ_i and $\upsilon_i^{'}$ are conditional, in that it is of the form $(\varepsilon_i \mid u_i > -z_i^{'}\gamma)$ for workers and $(\varepsilon_i \mid u_i \leq -z_i^{'}\gamma)$ for non-workers. Therefore, we can add variability to the predicted wage for workers and non-workers by drawing from the truncated bivariate normal distributions.[9] Drawing random error terms from these conditional distributions, we obtain predicted wages as follows:

$$\hat{\mathrm{w}} \mid_{E_i=1} = \mathrm{E}(\hat{\mathrm{w}}_i \mid_{E_i=1}) + (\hat{\varepsilon}_i \mid \hat{u}_i > -z_i^{'}\hat{\gamma}) \text{ for workers} \tag{13a}$$

$$\hat{\mathrm{w}} \mid_{E_i=0} = \mathrm{E}(\hat{\mathrm{w}}_i \mid_{E_i=0}) + (\hat{\varepsilon}_i \mid \hat{u}_i \leq -z_i^{'}\hat{\gamma}) \text{ for non-workers} \tag{13b}$$

where $E(\widehat{w}_i \mid_{E_i=1})$ *and* $E(\widehat{w}_i \mid_{E_i=0})$ are defined in (10a) and (10b).

For the unconditional wage, the error term is ε_i and we draw from $N(0, \hat{\sigma}_\varepsilon^2)$.

$$\hat{\mathrm{w}} = \mathrm{E}(\hat{\mathrm{w}}_i) + \hat{\varepsilon}_i \text{ for either workers or non-workers} \tag{14}$$

where $E(\widehat{w}_i)$ is defined in (9).

---

8 The variance of $\hat{\upsilon}_i$ is $\sigma_{\upsilon_i}^2 = \sigma_\varepsilon^2(1-\rho^2\delta_i)$ where $\delta_i = \lambda_i(\lambda_i + z_i'\gamma)$

9 We draw $\hat{\varepsilon}_i$ and $\hat{u}_i$ from the bivariate normal distribution $\begin{bmatrix} \hat{u}_i \\ \hat{\varepsilon}_i \end{bmatrix} \sim N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \hat{\rho}\hat{\sigma}_\varepsilon \\ \hat{\rho}\hat{\sigma}_\varepsilon & \hat{\sigma}_\varepsilon^2 \end{bmatrix}\right]$ and keep a draw $\hat{\varepsilon}_i$ if it meets the truncation condition $\hat{u}_i > -z_i'\hat{\gamma}$ for workers and $\hat{u}_i \leq -z_i'\hat{\gamma}$ for non-workers.

4 Data

The data are taken from the HES. The HES is produced by Statistics New Zealand and is conducted every year. Every 3 years HES collects detailed information on household expenditures and incomes and a range of demographic variables. Every other year, HES only collects income and demographic information. Households are interviewed throughout the year and the quarter of interview is recorded. On average, the sample is around 8,000 individuals in each survey year over the period 2006/07 to 2010/11.

We pooled the HES data of the 2006/07, 2007/08, 2008/09, 2009/10 and 2010/11 with a total sample size of 39,670 individuals. The sample is subdivided into five population groups (married men and women, single men and women and sole parents) and we omitted dependent children. Only 14% of sole parents are males and because of the small sample size of sole parents as a whole, both male and female sole parents are modelled as a single group.

We exclude individuals for whom we do not model their wages. Wages are modelled for the working age population (aged 16 to 64 inclusive), persons who are not on disability or sickness benefits, are not retired and are not full-time students. Since the wage equations in this paper are used in the labour supply models we do not model wages for persons engaged in any self-employment activity. This is due to the fact that for the self-employed there is a less direct link between the decision to work or to work an additional hour and the given wage rate than is the case for salary and wage workers (Kalb and Scutella 2003). Finally, we omit observations where wage rates are unrealistically small or large (less than 50% of the minimum wage relevant for the given period[10] or more than $150 per hour), where wage and salary or working-hours information do not match, and where industry and educational qualification information is missing. Omitting these individuals gives a final sample size of 17,639.

Sample means

The descriptions of all variables and the sample means of each variable are shown in Table 1. We applied similar variables in the wage and selection equations as previously in Kalb and Scutella (2003). The majority of variables are dummy variables with a value of 0 or 1. For example, *noqual* has a value of 1 if the person has less than school completion and a value of 0 if otherwise. A person is considered participating in work if they are currently earning some salary and wage income. The employment rate is given by *lpart* in Table 1.

The hourly wage rate is obtained by dividing the weekly salary and wage income from all current jobs by the number of weekly hours usually worked in all current jobs. Wage rates are adjusted to take account of wage growth over the period. They are adjusted by the quarterly male and female Average Weekly Earnings (AWE) index to the December quarter of 2011.[11]

Ethnicity is known to be an important factor affecting the offered wage rates and employment of individuals. Previous studies found a large variation in the average wages for different ethnicity groups. Maani (2000) found that Māori had lower qualifications than

[10] Minimum wage information is from the Ministry of Business, Innovation and Employment (2012).

[11] AWE data are from Statistics New Zealand (2006-2012).

non-Māori and that for given qualifications, their incomes were lower than non-Māori. On average, people from European descent tend to earn higher incomes compared to other ethnicities (Ministry of Education, 2008; Stillman, 2011). In addition, a quantitative survey of workforce non-participants aged 45 years and older found that people from the Māori, Pacific Islander and Asian ethnicities reported a greater number of job skill or workplace related barriers to enter paid employment than the New Zealand Europeans (Department of Labour, 2006).

The HES asks people to list multiple ethnicities. From this, we create several mutually exclusive ethnicity categories. If a person lists only one ethnicity and it is Māori or Pacific Islander, then they are categorised into *maoripaconly*. If they list more than one ethnicity and if one of them is Māori or Pacific Islander then they are categorised as *maoripacsome*. Of the remainder, if the person lists at least one ethnicity as European, then they are categorised as *europe*. The remainder are labelled *other*. Originally the *other* group was sub-divided into those that listed as non-Māori/Pacific and non-European ethnicity and those that listed several non-Māori/Pacific and non-European ethnicities. However, the sample size of the latter was very small when we ran the wage regressions and we decided to combine these into one category.

The household income variable includes only non-negative unearned private income; if there is negative income, the value of the variable is changed to zero. This income is adjusted by the Consumer Price Index (CPI) to make it comparable across the different HES datasets.[12] Tenure variables include an indicator of whether a person is living with his or her parents. This is only relevant to singles and sole parents. For persons who are in couple relationships, the partner's weekly earned income is included. This value is adjusted by male and female AWE index to make them comparable across the different HES datasets.

Industry categories are derived from the ANZSIC 2006 and occupational classification at the 1-digit level. For workers, the industry and occupational codes refer to the industry and occupation of the job in which the person works the most hours and each variable is assigned 0 or 1 depending on the industry and/or occupation of their main job. For non-workers, this information is unavailable. However, it is useful to include industry and occupation variables in the wage model as these typically have high explanatory power. One solution is to assign non-workers with the sample proportion of workers in the different industry and occupational categories rather than a value of 0 or 1. However the industry and occupational distribution of the unemployed is likely to be different for workers and non-workers (Creedy, Duncan, Harris and Scutella, 2001). We adopted the alternative used in Kalb and Scutella (2003) which uses a separate survey data source that reports the industry and occupation of the last job for people who are unemployed.[13] This information is used to assign a value to the industry variables that represents each industry's share of the total number of unemployed by gender. We do the same for the occupation variables of the non-workers.[14]

Wage profiles by age are typically an inverted "U" shape – wage rates usually increase with age and then drop for the oldest age groups. To capture this, we use the quadratic pair of age and age squared as regressors. Work experience is typically a strong determiner of the wage rate, however the HES does not contain a direct measure of work experience. One method is to assume experience to be the difference between age and

[12] CPI data are from Statistics New Zealand (2006-2012)

[13] The data are from the Household Labour Force Survey (HLFS) Statistics New Zealand (2006-2012)

[14] See Creedy *et al.* (2001) for further discussion of this approach.

the number of years spent in school and post-school study. However this would only be approximate as we do not have information about the number of years in school for people who do not complete high school and the years required for post-school qualifications. In addition, it is possible that individuals returning to study after some years in work rather than complete their studies before entering the workforce. An alternative, but still imperfect measure, is interacting education with age as indicated by the variables *pg_age*, *ba_age* and *voc_age*. These variables would also indirectly capture obsolescence of skills over time. We included other interaction variables *pg_mp*, *voc_mp*, *pg_ot* and *voc_ot* to capture any labour market disparity in relation to employment of skilled non-Europeans. Further descriptive statistics on employment rates and observed wages of the sample are in Appendix A.

The summary statistics in Table 1 are similar to the summary statistics in the previous study by Kalb and Scutella (2003). The employment rate is highest for married men and lowest for sole parents though the rates are higher than the previous study, indicating that a larger proportion of individuals are actively participating in the workforce. For married men and women this is also true for the employment rate for partners. The high rate of employment could be explained by the overall increase in education attainment observed in our sample. The proportions of individuals having no formal qualification have reduced significantly compared to the previous study. In addition, the majority of the demographic groups in our sample work in professional positions, with the exception of single men who mostly work in technical positions.[15] This is somewhat different from the previous study where most individuals were involved in non-professional occupations.

Employment rates

The employment rates by selected demographic variables are shown in Table A.1. The probability of working is highest for married men and lowest for female sole parents. The probability of working is higher for people with post-school qualifications than without, highest for Europeans, increases with age then drops again for the older age groups and is higher for households (except sole parents) that do not receive other non-labour private income. For women, the probability of working is lower if they have children, particularly young children, and if they have more children. For couples, the probability of working is also higher if the other partner is employed.

Observed wage rates

The wage rates by selected demographic variables are shown in Table A.2. The rates are observed to be the highest for married men, and as expected are highest for full-time workers compared to part-time workers, those with post-school qualifications compared to no qualifications, European compared to other ethnicities, living in Auckland and Wellington compared to the rest of the country, in managerial and professional occupations and tends to be highest in certain industries including the financial, business/real-estate, professional and scientific services, mining and public administration. Wage rates tend to be lowest for labourers and other lower skilled workers and in the retail trade and accommodation industries. The wage rate increases with age but then mostly decreases again for the highest age groups.

[15] This is consistent with the trend shown from the HLFS where majority of people in the survey are employed in professional positions and the proportion has grown consistently over time, from 2001 to 2011 (Statistics New Zealand, 2014).

Table 1 – Descriptive statistics: selected sample means

		Married men	Married women	Single men	Single women	Sole parents
No.observations		5676	6575	2180	2007	1201
No. employed		5083	4744	1813	1671	665
Variable	*Description*			*Sample mean*		
wage_ri (workers only)	observed wage (1)	31.66	26.35	23.09	23.33	24.42
lpart	participates in work	0.90	0.72	0.83	0.83	0.55
woman	woman					0.86
age10	age/10	4.34	4.24	3.41	3.86	3.74
kids	number of children	1.02	0.97			1.71
ageyk0	age of youngest child 0 years	0.08	0.08			0.07
ageyk1_3	age of youngest child 1 to 3 years	0.14	0.13			0.21
ageyk4_5	age of youngest child 4 to 5 years	0.05	0.05			0.11
ageyk6_9	age of youngest child 6 to 9 years	0.10	0.09			0.22
ageyk9_12	age of youngest child 10 to 12 years	0.07	0.07			0.16
ageykg12	age of youngest child over 12 years	0.10	0.10			0.24
Highest education qualification						
noqual	less than school qualification	0.13	0.12	0.17	0.13	0.20
school	completed school	0.27	0.35	0.38	0.36	0.41
bursary	bursary certificate	0.19	0.06	0.13	0.06	0.09
diploma	diploma (including vocational qualifications)	0.12	0.16	0.10	0.14	0.11
bachelor	bachelor degree	0.12	0.14	0.12	0.16	0.08
pgrad	post-graduate qualification	0.11	0.10	0.07	0.09	0.05
othqual	other qualification	0.05	0.06	0.04	0.06	0.05
Ethnicity						
europe	At least some European (but no Māori or Pacific Islander) descent	0.72	0.73	0.70	0.68	0.53
maoripaconly	Only of Māori or Pacific Islander descent	0.11	0.09	0.11	0.14	0.25
maoripacsome	Some Māori or Pacific Islander descent	0.04	0.04	0.07	0.06	0.11
other	Other non-European (but not Māori or Pacific Islander) descent	0.12	0.13	0.12	0.12	0.10
Interaction terms (residual group is non-postgraduate or non-vocational education and European)						
pg_mp	postgrad*(maoripaconly or maoripacsome)	0.01	0.02	0.01	0.02	0.01
voc_mp	(bursary or diploma or othqual)*(maoripaconly or maoripacsome)	0.05	0.03	0.04	0.04	0.07
pg_ot	postgrad*other	0.05	0.05	0.05	0.05	0.03
voc_ot	(bursary or diploma or othqual)*other	0.03	0.03	0.03	0.02	0.02
Region of residence						
nn_isl	Northern North Island	0.14	0.14	0.12	0.12	0.17
auckl	Auckland	0.27	0.28	0.28	0.29	0.30
cn_isl	Central North Island	0.17	0.17	0.17	0.17	0.14
well	Wellington	0.11	0.11	0.10	0.10	0.12
cant	Canterbury	0.17	0.16	0.18	0.18	0.16
sth_isl	South Island	0.15	0.15	0.14	0.13	0.11

		Married men	Married women	Single men	Single women	Sole parents
Industry of main job						
i_mine	Mining and agriculture	0.04	0.02	0.05	0.01	0.01
i_manu	manufacturing & electrical supply and related	0.21	0.08	0.19	0.07	0.11
i_cons	construction	0.12	0.02	0.12	0.01	0.04
i_wtrade	wholesale trade	0.07	0.04	0.07	0.04	0.04
i_rtrade	retail trade	0.06	0.11	0.1	0.12	0.09
i_accom	accommodation	0.02	0.04	0.06	0.08	0.07
i_trans	transport	0.06	0.03	0.06	0.03	0.03
i_comm	communications	0.02	0.02	0.02	0.03	0.02
i_fin	finance & business, real estate and related industries	0.04	0.06	0.03	0.06	0.04
i_scient	scientific, professional and related industries	0.08	0.07	0.07	0.07	0.06
i_public	public	0.1	0.09	0.07	0.09	0.07
i_educ	education	0.06	0.16	0.04	0.11	0.13
i_health	health	0.03	0.19	0.03	0.18	0.19
i_serv	services	0.07	0.08	0.1	0.09	0.1
Occupation of main job						
manag	manager	0.19	0.11	0.1	0.09	0.1
prof	professional	0.22	0.29	0.16	0.26	0.24
tech	technical	0.19	0.05	0.23	0.05	0.07
person	personal services and related	0.06	0.11	0.08	0.14	0.16
admin	administration and related	0.06	0.24	0.06	0.22	0.19
sales	sales	0.06	0.1	0.09	0.13	0.09
machin	machine operators	0.11	0.02	0.11	0.02	0.04
labour	labourers	0.11	0.07	0.17	0.08	0.11
Household variables						
hh/1000	non-labour private household income / 1000 (2)	3.23	4.39	0.85	1.47	1.35
maint	receives child maintenance payments	0.00	0.02		0.01	0.16
mort	owns own home with a mortgage	0.51	0.49	0.21	0.25	0.25
owner	owns own home outright	0.20	0.24	0.11	0.16	0.07
priv	private rental tenancy	0.24	0.22	0.35	0.33	0.48
publ	public rental tenancy	0.04	0.03	0.03	0.04	0.13
othtenure	other tenancy	0.01	0.01	0.01	0.01	0.01
livewp	lives with parents			0.29	0.20	0.06
un_rate	unemployment rate	5.05	5.43	5.00	5.44	5.28
yr	year trend	3.64	3.66	3.60	3.67	3.58
Partner variables						
pnoqual	less than school qualification	0.12	0.13			
pschool	completed school	0.35	0.27			
pbursary	bursary certificate	0.06	0.19			
pdiploma	diploma (including vocational qualifications)	0.15	0.13			
pbachelor	bachelor degree	0.14	0.12			
ppgrad	post-graduate qualification	0.11	0.11			

		Married men	Married women	Single men	Single women	Sole parents
pothqual	other qualification	0.06	0.05			
europepp	At least some European (but no Māori or Pacific Islander) descent	0.71	0.74			
maoripaconlyp	Only of Māori or Pacific Islander descent	0.10	0.10			
maoripacsomep	Some Māori or Pacific Islander descent	0.05	0.04			
otherp	Other non-European (but not Māori or Pacific Islander) descent	0.13	0.12			
partprt	whether partner is employed	0.73	0.88			
incprt/1000	partner weekly wage income / 1000	0.65	1.23			

Notes:

Observations are dropped from the sample for individuals who are (a) self-employed (b) disabled (c) on sickness benefits (d) full-time students (e) aged less than 15 years or over 64 years and (f) whose observed wage is less than half the minimum wage and more than $150 per hour.

(1) Observed wage is adjusted by the male and female average weekly earnings index to December quarter 2011. This figure is for workers only.

(2) Households with negative non-labour income are substituted with 0 non-labour income.

5 Results

In this section, we look at the main empirical results. Our results are set out in two sections. In the first section (5.1) we examine the coefficient estimates of the simultaneous models of selection and wage equations for each demographic group, focusing mainly on the significant variables. These are shown in Tables 2 to 4.[16] Next we consider the marginal effects of the estimated models. Overall, the correlation coefficient (*rho*) between the wage and selection equations is positive and significant for most groups, with the exception of the married women where the coefficient is not significant, and single men where the coefficient is negative and not significant. This indicates that sample selection bias exists for some groups and that correcting for this bias is justified.[17] The sizes of the standard errors (*sigma*) are similar, with small differences between all the demographic groups. The largest standard error is observed for sole parents while it is smallest for single women. A large standard error indicates that a large proportion of the differences in wage rates have not been explained by the variables in the model.

5.1 Coefficients

Couples

Overall, the coefficients[18] have expected signs in both the wage and selection equations for married men and women. For instance, the wage rates initially increase with age up to around 47 years for men and eventually decline with age.[19] The turning point is 49 years for married women. The previous study by Kalb and Scutella (2003) found turning points in the early forties. This increase could be explained by human capital theory and increasing education attainment especially for women over time in New Zealand. A large proportion of married women in our sample have diploma or bachelor degree qualifications and are engaged in professional occupations as opposed to lower educational qualifications and an elementary occupational level in the past (see Table 1). Human capital theory suggests that workers who invest more in education will also invest more in job training. Thus, the earnings of the better-educated workers rise more quickly because they are investing in job training and the rise will be for a longer time than their less-educated counterparts.[20]

As expected, there are considerable differences in wage rates between educational qualifications, occupations, industries, regions and ethnicity. Wage rates increase with the level of educational qualification, with the effect significant for married men. The interaction effect between education levels and age in the wage equation is positive and larger for people with higher education levels. This indicates that for persons with higher qualifications there is a more rapid increase in their wage rates as their age increase. This is consistent with Earle (2010), who found an effect of experience (measured by age) on wages for people with tertiary qualifications. Generally, wage rates are highest for

[16] We also estimated equations using the Heckman two-step approach, the results of which are available from the authors.

[17] We found similar results using the Heckman two-step approach.

[18] The interpretation of the effect of a one-unit change in a characteristic is calculated as follows: [exp(coefficient)-1]*100%. For example, a married man with a diploma will receive about 22% higher wages than a married man without qualifications.

[19] Non-linearity in earnings with age is also reported in Ball and Creedy (2013). However, in that study, the number of years in the labour force is used as a proxy for age, and annual income is used as a proxy for earnings.

[20] Another possible explanation for the turning points in age moving out over time is due to population ageing (increasing life expectancies and lower fertility rates).

managers and professionals are lowest for labourers. The finance sector seems to pay the highest wages while the hotel and accommodation sector pay the lowest wages. Married men living in Auckland and Wellington are paid higher wages than married men living elsewhere, with the lowest wages in Canterbury and South Island. We found that married women living in Wellington are paid lower than married women living in Auckland, and the largest wage gap is between genders in the South Island. Interestingly, married women are more likely to participate in the labour market if they live in Wellington and/or the South Island.

Married men and women from Māori and Pacific Islander descent and other non-European descent earn significantly less than people from European descent. The selection equation also shows that married men from the Māori and Pacific Islander group are less likely to participate in the workforce while married women from other non-European descent are less likely to participate. The interaction effect between ethnic group and education level in the wage equation indicates that both married women and men from other non-European descent benefit less from postgraduate education in so far as the wage rate they earn.

We included information on the partner's characteristics in the selection equation and found that a partner's employment has a positive effect on the person's employment for both married women and men, and is evidence of the complementarity of leisure (and work) between couples. Interestingly, the partner's wage income has a negative and significant effect on the person's employment for both genders, with the effect more pronounced for married men. The higher the partner's education level, married women are less likely to participate in the labour market. The effect is insignificant for men.

The selection equation of the model also includes some family compositional variables such as number of children, age of the youngest child and other non-labour household income. Similar to previous research, the presence of more children reduces labour force participation especially for married women. The presence of younger children has a larger impact than older children. These effects are not significant for married men. We also found that other non-labour household income reduces the probability of employment for both groups.

Singles

For singles, the wage rates initially increase with age and eventually decline with the turning point highest for single men (51 years). We found that the level of educational qualifications has no significant effect on wage rates and employment participation for single men. In contrast, the effect of education on wages is significant for single women with high school qualification and a bachelor degree. Single women with school qualifications and a bachelor degree are significantly more likely to participate in employment than a person with no formal qualifications. The results of interacting age with qualifications in the wage model show significant positive and larger effects for singles with higher education levels.

Similar to the results for married couples, the wage rates for singles working as managers and professionals are higher than singles working as labourers. The finance and public sectors seem to pay the highest wages for both genders. Single men working in the hotel and accommodation sector receive the lowest wages while single women receive the lowest wages in the retail sector.

Single women from Māori and Pacific Islander descent and other non-European descent earn significantly lower wages than their European counterparts. Singles from other non-European descent earn the lowest wage rates compared to other ethnicities. Interestingly,

this has a significant negative effect on the likelihood to participate in the labour market only for single women. Singles from Māori and Pacific Islander descent are less likely to participate in the workforce. The interaction effect between ethnic group and education level in both the wage and selection equations are insignificant for both genders, with the exception of other non-European single women with postgraduate qualifications. This indicates a weak positive effect of postgraduate qualification and the likelihood to participate in the labour market for other non-European single women.

While single women living in Wellington are more likely to participate in the workforce, they seem to receive slightly lower wages than their counterparts living in Auckland. The higher wage observed in Auckland is in line with those of Lewis and Stillman (2005), where they found that wage growth for wage and salary workers has been strongest in Auckland. The largest wage gap is observed for singles living in the north of the North Island compared to Auckland.

We also consider the unemployment rate and yearly time trend over the survey to capture the changes over time and the business cycle. The unemployment rate has a significant negative effect on the probability of employment for single men while it has no significant effect on single women's participation, and no effect on the wages of both genders. The yearly time trend is significant for single men indicating increased participation in labour market over the time of the survey. We also found that single men who live with their parents are less likely to participate in the labour market. In addition, single women who have other non-labour household income are found to be less likely to work.

Sole parents

Single mothers were found to earn significantly less and were less likely to work than single fathers; however the effect became insignificant once we controlled for occupational and industry categories as shown in Table 4. The presence of more children reduces labour force participation though it is not significant. Similar to married women, the presence of younger children has a larger negative impact than the presence of older children for sole parents' employment participation.

The age effect is weak for sole parents in both wage and selection equations. Sole parents with bachelor degree and postgraduate degree qualifications receive the highest wages and are more likely to work than those without formal qualifications. However, the interaction effect of the bachelor degree with age indicates a negative effect of qualification with age. Ethnicity does not have an effect on the wage and employment, with the exception of the other non-European group who have a significantly lower likelihood of working.

Sole parents working in the public and the scientific, professional and related sectors receive significantly higher wages than their counterparts in the services sector. The regional effect is weak in the wage equation but positively significant in the selection equation, specifically for sole parents living in the South Island and the central of North Island. Consistent with other groups, sole parents working in the managerial, professional and administrative occupation groups receive significantly higher wages than people working as labourers. Sole parents living with their parents are more likely to work possibly due to their parents providing childcare for them. However, this coefficient is insignificant in our model.

Table 2 – Maximum Likelihood estimates: Couples

	Wage equation (*lnwage*)		Selection equation (*lpart*)	
	Married men	Married women	Married men	Married women
age10	0.5245***	0.2751***	0.8473***	0.9537***
agesq	-0.0571***	-0.0297***	-0.1210***	-0.1298***
kids			-0.0032	-0.1342***
Age of youngest child (reference=no children)				
ageyk0			-0.0583	-1.3389***
ageyk1_3			-0.1428	-0.8970***
ageyk4_5			-0.1882	-0.5432***
ageyk6_9			-0.0536	-0.3089***
ageyk9_12			-0.1205	0.0393
ageykg12			0.062	0.0475
Education (reference=no qualifications)				
school	0.0438*	0.0459**	0.1279	0.2780***
bursary	0.1355**	0.0405	0.0987	0.5917**
diploma	0.2027***	0.0548	0.2706	0.6718***
bachelor	0.063	0.0834	0.6728	0.6821**
pgrad	0.0658	0.0883	0.0861	0.5753*
othqual	0.2285***	0.0364	0.2673	0.6498**
Ethnicity (reference=European)				
maoripaconly	-0.1618***	-0.1107***	-0.2823*	-0.1307
maoripacsome	-0.0439	-0.0518	-0.2839*	-0.2169*
other	-0.1008**	-0.1188***	-0.2607	-0.3255***
Interaction terms				
pg_age	0.0689***	0.0516***	0.0612	0.0395
ba_age	0.0525**	0.0311*	-0.0682	0.0045
voc_age	-0.0097	0.0145	0.0085	-0.0514
pg_mp	0.0246	-0.0038	0.284	0.3389
voc_mp	-0.0101	-0.0139	-0.0108	0.0522
pg_ot	-0.2103***	-0.0747*	0.0974	-0.3328**
voc_ot	-0.0233	0.0606	-0.0478	-0.195
Industry of current main job (reference=services)				
i_mine	-0.0411	-0.0515		
i_manu	0.1537***	0.0396		
i_cons	0.1134***	0.0867*		
i_wtrade	0.1236***	0.0801**		
i_rtrade	-0.0515	-0.1552***		
i_accom	-0.1877***	-0.1353***		
i_trans	0.1290***	0.0407		
i_comm	0.1181**	0.0877*		
i_fin	0.3633***	0.2032***		
i_scient	0.2789***	0.1484***		
i_public	0.2510***	0.1703***		
i_educ	-0.0375	-0.0793***		
i_health	0.0629	0.0108		

	Wage equation (*lnwage*)		Selection equation (*lpart*)	
	Married men	Married women	Married men	Married women
Occupation of current main job (reference=labourer)				
manag	0.3283***	0.4405***		
prof	0.2920***	0.3944***		
tech	0.1066***	0.1452***		
person	0.0321	0.0418		
admin	0.0825**	0.1859***		
sales	0.0770**	0.1028***		
machin	-0.0194	0.0149		
Region (reference=Auckland)				
nn_isl	-0.0860***	-0.1110***	-0.0232	-0.0111
well	0.0065	-0.0566***	0.0143	0.1832**
cn_isl	-0.0899***	-0.1167***	0.0459	0.0046
cant	-0.0989***	-0.0879***	0.1279	0.0422
sth_isl	-0.0959***	-0.1206***	-0.0219	0.1719**
Partner characteristics				
pschool			-0.0193	-0.1038
pbursary			-0.0332	-0.0443
pdiploma			0.0259	-0.1560*
pbachelor			0.011	-0.2159**
ppgrad			-0.1804	-0.3235***
pothqual			0.1076	-0.1256
maoripaconlyp			-0.0178	0.096
maoripacsomep			0.0727	0.0215
otherp			-0.1422	-0.025
incprt1000			-0.2047***	-0.0737***
partprt			0.5538***	0.5510***
Other variables				
un_rate	0.0146	-0.0004	-0.0112	0.0165
yr	-0.0083	-0.0017	0.0049	-0.006
hh1000			-0.0054***	-0.0050***
_cons	1.8218***	2.3223***	-0.1768	-1.1598***
Mills ratio and other statistics				
sigma $\hat{\sigma}_{\varepsilon}$	0.3599	0.3261		
rho $\hat{\rho}$	0.4785***	0.0474		
lambda/Inverse Mill's Ratio ($\hat{\rho}\hat{\sigma}_{\varepsilon}$)	0.1722***	0.0155		
Number of observations	5676	6575		
Number of censored observations	593	1831		
% correctly predicted (1)	89.7	75.8		
pseudo-R^2 (2)	0.29	0.27		

Notes: * p<0.05; ** p<0.01; *** p<0.001

Significance level on $\hat{\rho}$ represents the result of a Wald test $H_0 : \hat{\rho} = 0$

(1) % correctly predicted is a measure of the proportion of the sample whose labour force status is correctly predicted by the selection equation.

(2) $\text{Pseudo } R^2 \text{ is } 1 - \frac{\text{log likelihood/(N - number of parameters)}}{\text{restricted log likelihood/(N - 3)}}$. For details on the definition of restricted and unrestricted models see Kalb and Scutella (2003).

Table 3 – Maximum Likelihood estimates: Singles

	Wage equation (*lnwage*)		Selection equation (*lpart*)	
	Single men	Single women	Single men	Single women
age10	0.3695***	0.4444***	0.5779***	0.9784***
agesq	-0.0376***	-0.0513***	-0.0960***	-0.1295***
Education (reference=no qualifications)				
school	0.0435	0.0627*	0.1843	0.3998***
bursary	0.0923	0.0723	0.4312	0.3912
diploma	0.0681	0.1315*	0.3722	0.3064
bachelor	-0.0708	-0.0258	0.4806	1.1042**
pgrad	0.0446	-0.0971	0.6036	0.1573
othqual	0.0989	0.1082	0.3256	0.3572
Ethnicity (reference=European)				
maoripaconly	-0.0295	-0.1390***	-0.4803***	-0.6324***
maoripacsome	-0.0728*	-0.0383	-0.1473	-0.3951*
other	-0.1138**	-0.1593***	-0.2903	-0.6973***
Interaction terms				
pg_age	0.0577*	0.0705***	-0.0842	0.0591
ba_age	0.0766**	0.0326*	-0.0327	-0.1017
voc_age	0.001	-0.0066	-0.0177	0.0169
pg_mp	0.0507	0.0872	-0.378	0.4512
voc_mp	0.0459	0.0257	-0.0583	0.2447
pg_ot	0.0066	0.1037	-0.0527	0.5035*
voc_ot	-0.0082	-0.0946	0.0459	0.1552
Industry of current main job (reference=services)				
i_mine	0.008	0.0226		
i_manu	0.1137***	0.0836*		
i_cons	0.0239	0.0021		
i_wtrade	0.0552	0.0889*		
i_rtrade	-0.0499	-0.1250***		
i_accom	-0.0830*	-0.0529		
i_trans	0.0677	0.0668		
i_comm	0.1264*	-0.0012		
i_fin	0.2415***	0.1576***		
i_scient	0.1606***	0.1163***		
i_public	0.1897***	0.1532***		
i_educ	0.0711	0.0013		
i_health	0.0921	0.0750*		
Occupation of current main job (reference=labourer)				
manag	0.2579***	0.3968***		
prof	0.2812***	0.3869***		
tech	0.1339***	0.0945*		
person	0.0699	0.0449		
admin	0.0974**	0.2019***		
sales	0.0738*	0.1443***		
machin	0.0549	-0.0413		

	Wage equation (*lnwage*)		Selection equation (*lpart*)	
	Single men	Single women	Single men	Single women
Region (reference=Auckland)				
nn_isl	-0.0977***	-0.1476***	0.0719	-0.0302
well	-0.0179	-0.0578**	0.0558	0.2632*
cn_isl	-0.0849**	-0.1187***	0.0619	0.186
cant	-0.0638**	-0.1124***	0.2021	0.176
sth_isl	-0.0813**	-0.1223***	0.3056**	0.135
Other variables				
un_rate	0.013	-0.0003	-0.2159***	-0.12
yr	-0.0041	-0.0088	0.1137*	0.0215
hh1000			-0.0089	-0.0118***
livewp			-0.4193***	-0.0744
_cons	2.0617***	2.0079***	0.9424**	-0.2401
Mills ratio and other statistics				
sigma $\hat{\sigma}_{\varepsilon}$	0.310	0.2913		
rho $\hat{\rho}$	-0.2087	0.5561**		
lambda/Inverse Mill's Ratio ($\hat{\rho}\hat{\sigma}_{\varepsilon}$)	-0.0647	0.162***		
Number of observations	2180	2007		
Number of censored observations	367	336		
% correctly predicted (1)	82.8	83.1		
pseudo-R^2 (2)	0.27	0.37		

Notes: * p<0.05; ** p<0.01; *** p<0.001

Significance level on $\hat{\rho}$ represents the result of a Wald test $H_0 : \hat{\rho} = 0$

(1) % correctly predicted is a measure of the proportion of the sample whose labour force status is correctly predicted by the selection equation.

(2) $\text{Pseudo } R^2 \text{ is } 1 - \frac{\text{log likelihood/(N - number of parameters)}}{\text{restricted log likelihood/(N - 3)}}$. For details on the definition of restricted and unrestricted models see Kalb and Scutella (2003).

Table 4 – Maximum Likelihood estimates: Sole parents

	Wage equation (*lnwage*)	Selection equation (*lpart*)
age10	0.3247**	0.5454
agesq	-0.0285	-0.0536
kids		-0.0455
Age of youngest child (reference=children greater than 12 years)		
ageyk0		-0.9324***
ageyk1_3		-0.7366***
ageyk4_5		-0.4417**
ageyk6_9		-0.3969**
ageyk9_12		-0.176
Education (reference=no qualifications)		
school	0.1122*	0.4134***
bursary	0.1959	0.3371
diploma	0.2079	0.3069
bachelor	0.9269***	2.2529*
pgrad	0.8425*	4.1028*
othqual	0.1236	-0.1706
Ethnicity (reference=European)		
maoripaconly	-0.0402	-0.1544
maoripacsome	-0.0002	-0.1353
other	-0.0512	-0.3935*
Interaction terms		
pg_age	-0.063	-0.4819
ba_age	-0.1316*	-0.2894
voc_age	0.0148	0.0787
pg_mp	0.0507	0.1877
voc_mp	-0.1825*	-0.0191
pg_ot	-0.2683*	-0.6092
voc_ot	0.1205	0.3158
Industry of current main job (reference=services)		
i_mine	0.0568	
i_manu	0.1959***	
i_cons	0.2045*	
i_wtrade	0.1679*	
i_rtrade	-0.0051	
i_accom	-0.0568	
i_trans	0.2365**	
i_comm	0.2611**	
i_fin	0.2579***	
i_scient	0.2656***	
i_public	0.2940***	
i_educ	0.0606	
i_health	0.0854	

	Wage equation (*lnwage*)	Selection equation (*lpart*)
Occupation of current main job (reference=labourer)		
manag	0.4389***	
prof	0.3096***	
tech	0.028	
person	0.002	
admin	0.1914***	
sales	0.0092	
machin	-0.0356	
Region (reference=Auckland)		
nn_isl	-0.0694	0.1666
well	0.1018*	0.2138
cn_isl	0.008	0.3223*
cant	0.0266	0.2451
sth_isl	0.0651	0.4708***
Other variables		
un_rate	0.0008	-0.0255
yr	-0.0035	0.0204
hh1000		0.0027
livewp		0.0067
man	0.024	0.0954
_cons	1.6595***	-1.1322
Mills ratio and other statistics		
sigma $\hat{\sigma}_{\varepsilon}$	0.3649	
rho $\hat{\rho}$	0.6924**	
lambda/Inverse Mill's Ratio ($\hat{\rho}\hat{\sigma}_{\varepsilon}$)	0.2527***	
Number of observations	1201	
Number of censored observations	536	
% correctly predicted (1)	71.1	
pseudo-R^2 (2)	0.24	

Notes: * p<0.05; ** p<0.01; *** p<0.001

Significance level on $\hat{\rho}$ represents the result of a Wald test $H_0 : \hat{\rho} = 0$

(1) % correctly predicted is a measure of the proportion of the sample whose labour force status is correctly predicted by the selection equation.

(2) $\text{Pseudo } R^2 \text{ is } 1 - \frac{\text{log likelihood/(N - number of parameters)}}{\text{restricted log likelihood/(N - 3)}}$. For details on the definition of restricted and unrestricted models see Kalb and Scutella (2003).

5.2 Marginal effects

In this subsection, we provide further interpretation of the wage differences between groups with respect to an observed characteristic. We follow similar examples used previously in Kalb and Scutella (2003) and report on the differences with our results.

We first consider the impact of postgraduate qualifications on the wages of all demographic groups of European descent. We found that single men and married women who are 30 years old with postgraduate qualifications and from European descent are expected to receive about 25% higher wages than those without formal qualifications.[21] Married men with similar qualifications are expected to receive a wage rate which is about 31% higher than those without formal qualifications, while single women receive a wage 12% higher. Interestingly, a sole parent of 30 years with a postgraduate degree is expected to receive a wage of 92% higher than those without formal qualifications. Our results indicate that sole parents experience the largest effect from higher education on wage levels, thus leading to the increased likelihood to participate in employment. Generally, the wage gap between a person with postgraduate and without formal qualifications has narrowed for most groups compared to the previous study, except for married men and sole parents. On average, for all groups except for sole parents, individuals with bachelor degree qualifications have wage rates which are between 8 and 25% higher than those without formal qualifications. The highest effect is experienced by sole parents (70% higher).

Next, we consider the regional impact on wages for all groups. Specifically we explore the impact of living in the north of the North Island where the wage rates are lower for all groups compared to their counterparts living in Auckland. On average, all groups living in the north of the North Island receive wages which are between 8 and 14% lower than in Auckland, with sole parents experiencing the smallest effect of 7%.

Lastly, we consider the impact of age on the wages of all demographic groups. We found that the wages for married men with no formal qualifications increase 20% for a ten-year increase in age from 25 to 35 years and a 7% increase for a ten-year increase from 35 to 45 years of age. Our results show a larger wage increase in the ten-year age gap compared to the previous study. Our results also show that the turning point where wage rates start to decrease with age happen later in the mid forties or early fifties of an individuals' life, as opposed to early forties in the previous study. We further analyse the wages of married men with postgraduate qualifications for a ten-year increase in age. The wages increase 29% and 15% for a ten-year increase in age from 25 to 35 years and 35 to 45 years of age respectively. Consistent with our earlier results and previous findings, the age-earnings profiles are steeper for higher educated people and these people tend to receive their maximum wage rate at an older age compared to people with no qualifications.

[21] See footnote 19 on the calculation of the marginal effects.

6 Wage predictions

Conditional wage predictions using equation (13a) for workers and equation (13b) for non-workers are shown in Table 5. The table shows the average predictions by labour force status and education level. These predictions are generated using the conditional mean wage (equations 10a and 10b) and adding a random disturbance term. For all groups, predicted and observed wages are numerically close but are nonetheless significantly different.[22]

As expected, with the exception of single men, workers have significantly higher average predicted wages than non-workers. Workers also have significantly higher average predicted wages than non-workers for each education level.

The difference in single men's predicted wages between workers and non-workers is driven by the estimated negative sample selection term (represented by the covariance $\sigma_{12} = \rho\sigma_{\varepsilon}$ in equations (13a) and (13b)) for this group. For single men, workers have lower average predicted wages compared to non-workers for the *same* education level, but this is significant only for the first two education levels. However, given *different* education levels, working single men with bachelor and post-graduate education have higher average predicted wages than non-workers with lower education levels as we would expect.

For the other population groups, positive σ_{12} terms indicate that sample selection is such that the characteristics of the employed and not employed are different and that these differences are reflected in higher offered wage rates for workers. However, as suggested in Breunig and Mercante (2010), the sample selection term in selection models captures two effects. Firstly, the unobservable characteristics that result in a higher employment probability also result in a higher wage. This is what we observe for all groups other than single men who have a negative but insignificant sample selection term. Secondly, the sample selection term captures the difference between the variance of wage offers and the covariance between wage offers and reservation wages. The sample selection term could be negative if the covariance between reservation wages and wage offers (which we expect to be positive) is greater than the variance of wage offers. Therefore a negative sample selection term does not imply that differences in characteristics of the employed and not employed are reflected in lower offered wage rates for workers compared to non-workers.[23]

For wage predictions however, it seems counter-intuitive to predict higher wages for non-workers compared to workers. An alternative to the conditional wage is to use the linear predicted wage.[24] In a study that compared the performance of wage predictors for non-workers from sample selection models, it was found that the linear predictor often

[22] As a comparison, the conditional mean wage predictions using equations (10a) and (10b) are shown in Table A.3 in Appendix A. For these predictions, we find the difference between the observed and predicted wages of workers are largely not significant. The main reason for the difference between the conditional mean wage and the conditional wage with a random term is the inclusion of variability in the prediction by the extra random term.

[23] The issue of a negative sample selection term is discussed at length in Ermisch and Wright (1994), who argue that in a "reservation model of labour supply" (which we are implicitly using) a negative sample selection term is plausible and that a negative selection bias is not necessarily a cause for concern.

[24] Linear estimates are obtained from Equation (4) over the sample of workers using ordinary least squares. Essentially it assumes u_i and ε_i are independent. The linear coefficient estimates are in appendix Table A.4. Yet another alternative is to use the conditional wage predictor (equation (14)). We do not show these results; however they are available from the authors.

outperforms the conditional predictor (Breunig and Mercante (2010)), largely driven by instability and imprecision in the estimated sample selection term.[25] Using the linear predictor is also plausible because the sample selection term $\sigma_{12} = \rho\sigma_{\varepsilon}$ is not significant for single men. Using the linear model the predicted wages of workers were found to be higher than for non-workers; however the differences were not statistically significant.[26]

The predicted conditional wage distribution for sole parents is shown in Figures B.1 to B.4 in Appendix B for both log and level forms. The figures show visually how adding additional variability to the mean wage prediction widens the predicted wage distributions of workers and non-workers. This widening of the predicted wage distribution is also found in the other population groups.

A wider distribution approximates more closely the observed wage distribution. However, a wider distribution may result in a greater number of predicted wages falling below minimum wage levels, particularly for non-workers. For sole parents, for example, 3.3% of workers had observed wages less than the minimum wage.[27] About 30% of non-workers have a mean predicted wage rate below the minimum wage, and when adding variability through a random disturbance term, around 60% have a predicted wage rate below the minimum wage. A similar situation occurs with single women and married men, though for these groups the proportions below the minimum wage is smaller compared to sole parents. This is not surprising given that all three groups were found to have significant sample selection terms. Of the three, sole parents had a larger estimated variance of the error term ($\hat{\sigma}_{\varepsilon}$) giving rise to greater variability in the added random disturbance terms.

In general, wage predictions using the estimated models follow patterns consistent with the underlying data (see Table 5). For all groups, higher educational qualifications are reflected in higher wage rates. For example, working married men with a bachelor degree have an average wage rate of around $16 per hour higher than those not having completed school. However, for non-working married men, the gap is around $5 for the two education groups. We further divided the predicted wages by other characteristics and found that the Europeans have higher average wage rates and Maori and Pacific Islanders the lowest. Overall, the predicted average wages are higher in Auckland and Wellington compared to the rest of the country. The predicted average wages are highest in the public, finance and business and professional and scientific services sectors and are lowest in the accommodation and food and retail sectors. Professionals and managerial occupations have higher wage rates compared to lower skilled occupations such as labourers, machine operators, sales and personal service occupations. The predicted wages increase with age, reach a peak and then generally drop for the oldest age groups.

Wage predictions are the result of the interplay of a number of characteristics, some significant and others not, reflected in the estimated coefficients of the wage equations. In Table 6 we show predicted wages for selected hypothetical cases. The table shows the conditional mean wage (column 6) and the predicted conditional wage with a random disturbance term (column 7). Column 7 includes an indication of the confidence band of the prediction. We see the large impact of education on wages: for example a professional working married woman with a bachelor degree has an expected conditional wage rate of

---

25 That study reported Heckman two-step estimates of the sample selection model; however the results would also apply to maximum likelihood estimates of the sample selection model.

26 The linear wage predictions for single men are shown in Appendix A.5.

27 Note that workers with wage rates less than 50% of the minimum wage rate were omitted from the sample. However, only a very small number of observations were omitted for this reason.

$41 per hour compared to another married woman working as a clerk who has completed school with a wage rate of $29 (column 7 for cases 1 and 2). However, a married woman with no formal qualifications and working in retail trade results in a lower wage rate of $18 per hour (case 3). This is also true if we compare the wages predicted for a professional working single man with a bachelor degree and a non-working single man with no formal qualifications, working in the same industry (case 5). In most cases where the individual has the same qualifications (with the exception of single man in case 6), wage offers for workers are predicted to be higher than for non-workers.

The table shows the impact of the negative sample selection term on a single male worker in manufacturing who has an offered wage rate lower than a non-worker with the *same* characteristics (case 6). If instead we were to use the linear predicted wage there is hardly any difference in the wage predictions of the hypothetical worker compared to the non-worker.[28] The table also shows the large prediction interval when using predicted wages with a random component (column 7). Standard deviations of the predicted wages with a random term are quite large.

[28] This is shown in the appendix Table A.6.

Table 5 – Average predicted conditional hourly wage rates with random disturbance term by labour force status and education ($ per hour) (Equations 13a and 13b)

	Married men	Married women	Single men	Single women	Sole parents (2)
Non-workers	19.46	21.27	25.09	14.73	12.95 (13.3,12.9)
Workers (1)	31.22	26.13	22.98	23.04	24.08 (26.6, 23.6)
	31.66	*26.35*	*23.09*	*23.33*	*24.42 (26.87, 23.94)*
Non-workers					
noqual	18.29	18.51	24.64	13.69	12.32
school	17.36	20.38	23.03	14.75	12.89
bursary	18.68	20.11	25.7†	13.85	12.56
diploma	20.41	22.78	25.24†	15.06	13.9
bachelor	22.94	23.33	29.33†	15.73	14.57
pgrad	24.0	26.22	33.58†	17.82	17.11‡
othqual	21.57	21.27	23.54†	14.53	13.76
Workers (1)					
noqual	23.19	19.96	19.74	18.43	19.61
	23.1	*19.6*	*19.5*	*18.6*	*19.3*
school	27.04	23.03	20.53	20.84	21.59
	26.9	*22.9*	*20.0*	*20.7*	*20.9*
bursary	28.09	24.79	22.69†	21.94	23.33
	28.1	*24.3*	*23.4*	*21.5*	*22.5*
diploma	33.61	27.03	23.8†	24.16	24.69
	33.5	*27.5*	*23.8*	*25.2*	*26.0*
bachelor	39.35	31.14	28.3†	25.18	30.83
	40.1	*31.4*	*28.9*	*26.0*	*30.2*
pgrad	43.84	35.02	32.89†	31.70	33.91
	45.9	*36.6*	*34.4*	*32.4*	*39.7*
othqual	32.12	24.82	23.26†	23.43	24.14
	35.6	*25.3*	*25.3*	*23.4*	*26.8*

Notes:

(1) Observed wages for workers are shown in italics.

(2) Sole parent male and female wage rates are in parentheses.

‡ This value is based on fewer than 5 individuals.

Wage rates between workers and non-workers overall and between workers and non-workers for each education level are significantly different at 95% confidence level, except where indicated by †.

Table 6 – Wage predictions for hypothetical individuals ($ per hour)

	case	status	education	Industry	conditional mean wage eqn (10a & b)	predicted conditional wage (with random term) (1) eqn (13a & b)
1	married woman	not working	bachelor	professional services	39.4	39.8 (23.1, 65.1)
		working	bachelor	professional services	40.8	40.6 (22.6, 65.5)
2	married woman	not working	school completed	public sector	28.5	28.7 (14.9, 46.4)
		working	school completed	public sector	29.2	29.1 (15.9, 47)
3	married woman	not working	no qualification	retail trade	18.1	18.2 (10.3, 29.4)
		working	no qualification	retail trade	18.5	18.3 (10.3, 29.9)
4	sole parent	not working	no qualification	finance	17.0	17.0 (9.3, 27.7)
		working	no qualification	finance	26.1	26.2 (16.0, 40.1)
5	single man	not working	noqual	professional services	34.6	34.5 (20.6, 54.2)
		working	bachelor	professional services	37.6	37.6 (22, 58.1)
6	single man	not working	school completed	manufacturing	27.8	28.4 (16.6, 45.5)
		working	school completed	manufacturing	24.5	24.4 (14, 39.2)

Notes:

(1) The estimates are the average of 1000 draws of the predicted conditional wage with a random term. The 90% prediction interval over the 1000 draws is shown in parentheses.

(2) All cases are for individuals 35 years old, living in Auckland, European and no other household income. Married women and sole parent cases considered have 1 child who is 3 years old. Married women have a working partner.

(3) Estimated wage rates are adjusted by Average Weekly Earnings to December quarter 2011.

7 Conclusion

In this paper we estimate wage equations for five different population groups in New Zealand using the HES over the years 2006/07 to 2010/11. We update previous estimates by Kalb and Scutella (2003) using the HES over the period 1991/92 to 2000/01. The possibility of sample selection bias is taken into account by using a sample selection model. We estimate wages using maximum likelihood and the Heckman two-step procedure, though we report only the former.

In estimating the wage equations we take account of education, ethnicity, geography and industry and occupation. We also take account of the trend in wages over time and the business cycle by including controls for the unemployment rate and a time trend variable. In controlling for sample selection we take account of other characteristics such as the presence and ages of children, the presence of unearned income and for couples, whether the partner works and other selected characteristics of the partner.

The sample selection is positive and significant for married men, single women and sole parents. It is positive and not significant for married women. It is negative for single men, though not significant. In the previous study, sample selection was found to be positive and significant for all population groups except sole parents, where the term was negative but not significant. We found that the wage coefficient estimates generally agree with those of Kalb and Scutella (2003), however there are some important differences.

Overall, wage rates are positively related to age and education for all demographic groups. Generally, people at higher education levels experience a steeper wage increase and reach the maximum rate at an older age. The turning points where wage rates start to decrease with age occur later in the mid forties or early fifties of an individual's life, as opposed to early forties from the past research. This could be explained by the human capital theory and increasing educational attainment over time in New Zealand. The educational attainment has improved over the period, especially for women. In addition, the human capital theory suggests that workers who invested more in education will also invest more in job training. Thus, the earnings of the better-educated workers rise more quickly because they are investing in job training and the rise will be for a longer time than their less-educated counterparts.

There are considerable differences in wage rates between occupations, industries, regions and ethnicity for all demographic groups. People working in managerial and professional positions experience higher wages than people working in elementary occupations. People working in the finance and public sectors are paid higher wages than the services sector and people living in Auckland are the highest wage earners. Individuals from European descent are paid more than individuals from other ethnicities, though this is not significant for sole parents.

The presence of more children significantly reduces the probability of married women's participation in employment. The presence of younger children has a large impact on labour force participation for both married women and sole parents. The partner's wage income has a negative and significant effect on the person's employment for both married couples. Employment participation is significantly higher for married women, single women and sole parents with higher education, from European descent, and residing in Wellington and South Island. Women married to a partner with higher education are less likely to participate in the labour market.

We used the estimated wage equations to predict hourly wage rates for individuals in our sample. In predicting wages of workers and non-workers, we found that the mean wage has a narrower distribution than the distribution of observed wages. We additionally predicted wages with a random disturbance term to introduce a wider dispersion in the wage distribution. This is particularly useful in labour supply simulation models which typically require wage rates to be imputed for non-workers. However, the greater dispersion of the predicted wage distribution can result in wages below the statutory minimum wage. We found this to be the case particularly for sole parents.

We found that the wage predictions generally follow patterns consistent with the underlying data. However, the negative sample selection for single men, results in a number of cases of hourly wage predictions of workers to be smaller than the predictions of non-workers. For simulation models, an alternative could be to use the linear predictor when there is no evidence of sample selection bias as was the case with single men. Additionally, the linear wage predictor has previously been shown to often be a better predictor of the wage rate of non-workers.

The estimated equations are used in Treasury's behavioural microsimulation model, (TAXMOD-B) to predict labour supply responses to policy changes. For the behavioural microsimulation model to produce reliable results, it is helpful that the wage predictions are based on the most up-to-date data.

References

Amemiya, Takeshi (1985) *Advanced econometrics.* (Oxford: Basil Blackwell).

Ball, Christopher and John Creedy (2013) "Population ageing and the growth of income and consumption tax revenue." Wellington, New Zealand Treasury Working Paper 13/09, April.

Breunig, Robert and Joseph Mercante (2010) "The accuracy of predicted wages of the non-employed and implications for policy simulations from structural labour supply models." *The Economic Record* 86(272): 49-70.

Cameron, A Colin and Pravin K Trivedi (2005) *Microeconometrics: Methods and applications*. (New York: Cambridge University Press).

Creedy, John, Alan S Duncan, Mark N Harris and Rosanna Scutella (2001) "Wage functions for demographic groups in Australia." *Australian Journal of Labour Economics* 4(4): 296-316.

Dixon, Sylvia (2000) "Pay inequality between men and women in New Zealand." Wellington, Department of Labour Occasional Paper 1, September.

Department of Labour (2006) "45 Plus: Choices in the labour market: Quantitative survey." Wellington, Department of Labour, November. <http://www.dol.govt.nz/publications/research/45plus-quantitative/>

Earle, David (2010) "Labour market outcomes of skills and qualifications." Wellington, The Tertiary Education Occasional Paper, Ministry of Education, March. <http://www.educationcounts.govt.nz>

Ermisch, John F and Robert E Wright (1994) "Interpretation of negative sample selection effects in wage offer equations." *Applied Economics Letters* 1(11): 187-189.

Greene, William H (2003) *Econometric analysis*, Fifth edition. (New Jersey:Prentice Hall).

Heckman, James J (1979) "Sample selection bias as a specification error." *Econometrica* 47(1): 153-161.

Hyslop, Dean and David C Mare (2009) "Skill upgrading in New Zealand, 1986-2001." *Australian Economic Review* 42(4): 422-434.

Kalb, Guyonne and Rosanna Scutella (2003) "Wage and employment rates in New Zealand from 1991 to 2001." Wellington, New Zealand Treasury Working Paper 03/13, June.

Lewis, Geoff and Steve Stillman (2005) "Regional economic performance in New Zealand: How does Auckland compare?" Wellington, New Zealand Treasury Working Paper 05/08, November.

Maani, S. (1999) "Private and Public Returns to Investments in Secondary and Higher Education in New Zealand Over Time: 1981–1996." Wellington, New Zealand Treasury Working Paper 99/02, March.

Maani, S. (2000) “Secondary and tertiary educational attainment and income levels for Maori and non-Maori over time.” Wellington, New Zealand Treasury Working Paper 00/18,December.

Ministry of Business, Innovation and Employment (2012) “Previous minimum wage rates.” Wellington, Ministry of Business, Innovation and Employment. <http://www.dol.govt.nz/er/pay/minimumwage/previousminimum.asp>

Ministry of Education (2008) “State of education in New Zealand: 2008.” Wellington, Ministry of Education, December. <http://www.educationcounts.govt.nz/publications/ece/2551/34702/part-4---labour-market>

Ministry of Education (2011) “Impact of education on income (factsheet).” Wellington, Ministry of Education, August. <http://*www.educationcounts.govt.nz/indicators/main/education-and-learning-outcomes/1919*>

Ministry of Education (2012) “Profile and trends 2011: New Zealand’s Tertiary Education Sector.” Wellington, Ministry of Education, November. <http://www.educationcounts.govt.nz/publications/series/2531/114396>

Ministry of Women’s Affair (2013). <http://mwa.govt.nz/gender-pay-gap-0>

Smart, Warren (2006) “Outcomes of the New Zealand tertiary education system: A synthesis of the evidence.” Wellington, Ministry of Education, August.

Statistics New Zealand (2006-2012) Consumers Price Index (CPI).

Statistics New Zealand (2006-2012) Earnings and Employment Survey (QES).

Statistics New Zealand (2006-2011) Household Economic Survey (HES).

Statistics New Zealand (2006-2012) Household Labour Force Survey (HLFS).

Stillman, Steve (2011) “Labour market outcomes for immigrants and the New Zealand-born 1997-2009.” Wellington, Department of Labour, May. <http://dol.govt.nz/publications/research/labour-market-outcomes/index.asp>

Appendix A – Additional tables

Table A.1 – Descriptive statistics: employment rates (%)

	Married men	Married women	Single men	Single women	Sole parents
All persons	89.6	72.2	83.2	83.3	55.4
men					65.5
women					53.8
Education					
noqual	85.1	63.4	75.6	67.9	37.1
school	89.1	70.1	82.6	82.2	51.0
bursary	89.6	73.5	86.4	87.5	66.7
diploma	91.9	74.8	86.5	85.0	67.4
bachelor	92.3	75.8	88.1	92.5	73.0
pgrad	90.3	79.3	84.5	88.1	94.7
othqual	89.7	73.4	85.2	83.8	50.8
Ethnicity					
europe	90.8	75.6	85.9	87.3	64.4
maoripaconly	86.9	65.5	70.0	69.7	42.4
maoripacsome	86.9	67.6	81.0	81.2	45.3
othonly	85.7	58.9	81.0	76.5	50.8
Region					
nn_isl	88.5	70.0	80.7	77.2	51.0
auckl	88.2	68.1	80.1	79.5	48.7
well	90.2	71.8	81.9	83.8	58.0
cn_isl	89.9	75.7	83.2	87.1	59.1
cant	91.9	72.8	87.1	87.5	59.5
sth_isl	89.4	77.1	87.3	85.8	65.9
Tenure					
mort	93.3	78.3	90.1	88.4	78.1
own	82.2	69.7	71.9	75.6	74.4
priv	89.2	64.9	88.3	85.8	49.0
publ	82.1	51.5	68.1	74.4	34.2
other	87.0	61.8	75.0	70.8	50.0
livewp			78.1	81.5	37.7
Partner LFS					
partner works	92.0	74.5			
partner does not work	83.1	55.0			
Other household income					
receives other household income	87.0	72.7	81.3	80.7	71.6
does not receives other household income	90.4	72.0	83.5	83.8	52.7

	Married men	Married women	Single men	Single women	Sole parents
Maintenance					
receives maintenance					85.7
does not receive maintenance					49.7
Age youngest child					
none	87.8	79.7			
ageyk0	90.8	37.5			22.8
ageyk1_3	90.8	53.8			34.3
ageyk4_5	90.5	65.2			50.4
ageyk6_9	92.3	72.8			55.8
ageyk9_12	90.3	81.8			65.6
ageykg12	91.3	81.6			78.4
Kids					
0	87.8	79.7			
1	90.7	68.7			60.1
2	91.5	66.2			55.2
3 or more	91.0	55.2			40.7

Table A.2 – Descriptive statistics: observed wages ($ per hour)

	Married men	Married women	Single men	Single women	Sole parents
Observed wage rate	31.7	26.3	23.1	23.3	24.4
Labour force status					
Full-time	31.9	27.3	23.5	24.3	27.0
Part-time	27.2	24.6	19.0	19.2	20.6
Education					
noqual	23.1	19.6	19.5	18.6	19.3
school	26.9	22.9	20.0	20.7	20.9
bursary	28.1	24.3	23.4	21.5	22.5
diploma	33.5	27.5	23.8	25.2	26.0
bachelor	40.1	31.4	28.9	26.0	30.2
pgrad	45.9	36.6	34.4	32.4	39.7
othqual	35.6	25.3	25.3	23.4	26.8
Ethnicity					
europe	33.6	27.3	24.1	24.1	25.8
maoripaconly	23.5	21.5	20.6	20.3	21.4
maoripacsome	27.2	24.3	19.0	21.4	21.6
othonly	28.9	23.9	21.1	22.9	24.0
Region					
nn_isl	28.7	24.1	19.7	20.0	21.1
auckl	33.8	28.4	24.4	25.8	25.6
well	27.8	24.4	21.8	21.4	21.8
cn_isl	38.1	29.4	26.2	25.7	30.6
cant	29.1	24.9	22.1	21.6	23.5
sth_isl	28.9	24.2	22.0	21.6	23.0
Industry					
i_agr	21.7	21.1	18.9	16.9	16.5
i_mine	37.4	35.5	32.8		41.3
i_manu	28.4	22.8	22.4	20.6	23.9
i_elec	36.3	30.8	25.5	24.5	27.4
i_cons	27.5	26.9	20.7	21.1	25.7
i_wtrade	31.5	27.4	21.2	26.4	23.5
i_rtrade	24.5	19.3	18.3	17.2	17.6
i_accom	20.3	18.5	16.6	16.0	15.7
i_trans	28.5	24.3	22.2	22.4	23.2
i_comm	35.4	31.1	27.9	24.1	28.5
i_fin	51.2	32.3	37.8	29.3	30.4
i_bus	42.5	28.1	23.7	21.4	19.2
i_scient	45.4	32.6	31.2	27.4	35.2
i_serv	28.8	24.6	21.3	21.4	19.6
i_public	36.4	32.0	28.9	28.7	32.6
i_educ	32.9	27.7	29.7	26.8	28.4
i_health	34.9	26.4	26.8	24.9	23.5

	Married men	Married women	Single men	Single women	Sole parents
Occupation					
manag	40.9	34.5	28.1	29.5	36.2
prof	40.4	32.9	31.9	30.1	31.7
tech	26.8	22.6	21.5	19.0	21.4
person	27.2	19.6	22.0	18.1	18.5
admin	29.9	24.7	23.4	23.5	23.6
sales	26.5	20.2	19.7	18.2	17.6
machin	22.6	18.8	20.1	16.6	18.0
labour	21.5	17.3	17.8	16.7	17.1

Table A.3 –Average of the mean conditional hourly wage rates by labour force status and education without random term ($ per hour) (Equations 10a and 10b)

	Married men	Married women	Single men	Single women	Sole parents (2)
Non-workers	19.83	21.45	24.45	14.89	12.82 (13.42, 12.75)
Workers (1)	31.4	26.14	22.91	23.21	24.11 (26.59, 23.63)
	31.66	*26.35*	*23.09*	*23.33*	*24.42 (26.87, 23.94)*
Non-workers					
noqual	17.26	19.24	22.35	14.34	12.03
school	18.38	20.3	22.75	14.55	12.74
bursary	19.34	21.42	26.06	14.02	12.81
diploma	20.24	22.53	26.07	15.37	14.05
bachelor	23.49	23.71	28.32†	15.98	14.3
pgrad	25.25	26.01	32.32†	18.04	14.8‡
othqual	21.37	21.8	26.36	14.57	14.03
Workers (1)					
noqual	23.41	19.77	19.76	18.81	18.71
	23.1	*19.6*	*19.5*	*18.6*	*19.3*
school	27.01	23.09	20.06	20.8	21.19
	26.9	*22.9*	*20.0*	*20.7*	*20.9*
bursary	28.53	24.35	23.51	21.94	23.22
	28.1	*24.3*	*23.4*	*21.5*	*22.5*
diploma	33.4	26.98	23.76	24.66	25.43
	33.5	*27.5*	*23.8*	*25.2*	*26.0*
bachelor	39.15	30.97	28.29†	25.7	29.79
	40.1	*31.4*	*28.9*	*26.0*	*30.2*
pgrad	44.63	35.69	32.59†	31.75	37.23
	45.9	*36.6*	*34.4*	*32.4*	*39.7*
othqual	32.88	24.79	23.95	23.08	25.44
	35.6	*25.3*	*25.3*	*23.4*	*26.8*

Notes:

(1) Observed wages for workers are shown in italics.

(2) Sole parent male and female wage rates are in parentheses.

‡ This value is based on fewer than 5 individuals.

The wage rates between workers and non-workers overall and between workers and non-workers for each education level are significantly different at 95% confidence level, except where indicated by †.

Table A.4 – Linear estimates: Single men

age10	0.3901***
agesq	-0.0406***
Education (reference=no qualifications)	
school	0.0482*
bursary	0.1023
diploma	0.0765
bachelor	-0.0598
pgrad	0.063
othqual	0.1059
Ethnicity (reference=European)	
maoripaconly	-0.0422
maoripacsome	-0.0762*
other	-0.1207**
Interaction terms	
pg_age	0.0544*
ba_age	0.0760**
voc_age	0.0008
pg_mp	0.0435
voc_mp	0.0457
pg_ot	0.0068
voc_ot	-0.0058
Industry of current main job (reference=services)	
i_mine	0.0082
i_manu	0.1134***
i_cons	0.0238
i_wtrade	0.0546
i_rtrade	-0.0501
i_accom	-0.0820*
i_trans	0.068
i_comm	0.1268*
i_fin	0.2426***
i_scient	0.1602***
i_public	0.1902***
i_educ	0.0707
i_health	0.0913
Occupation of current main job (reference=labourer)	
manag	0.2583***
prof	0.2813***
tech	0.1339***
person	0.0701
admin	0.0969*
sales	0.0739*
machin	0.0547

	Single men
Region (reference=Auckland)	
nn_isl	-0.0953***
well	-0.016
cn_isl	-0.0828**
cant	-0.0584*
sth_isl	-0.0741**
Other variables	
un_rate	0.0078
yr	-0.0013
_cons	2.0233***
Other statistics	
Number of observations	1810
adjusted-R^2	0.4

Notes: * p<0.05; ** p<0.01; *** p<0.001

Table A.5 – Average of the predicted hourly wage rates of single men by labour force status and education using the linear predictor ($ per hour)

	Single men
Non-workers	21.99
Workers (1)	22.81
	23.09
Non-workers	
noqual	20.15
school	20.51
bursary	22.28
diploma	22.99
bachelor	25.43
pgrad	28.89
othqual	27.98
Workers (1)	
noqual	20.03
	19.5
school	20.04
	20.0
bursary	23.29
	23.4
diploma	23.97
	23.8
bachelor	27.48
	28.9
pgrad	31.97
	34.4
othqual	24.60
	25.3

Notes:

(1) Observed wages for workers are shown in italics.

Wage rates between workers and non-workers overall and between workers and non-workers for each education level were not significantly different at 95% confidence level except where indicated by a †.

Table A.6 – Wage predictions for hypothetical single men using a linear predictor ($ per hour)

	case	status	education	Industry	mean wage eqn (4)	predicted wage (with random term) (1)
5	single man	not working	noqual	professional services	30.6	30.5 (17.6, 48.8)
		working	bachelor	professional services	37.6	37.4 (21.5, 59.3)
6	single man	not working	school completed	manufacturing	24.4	24.1 (14.2, 37.8)
		working	school completed	manufacturing	24.4	23.9 (13.8, 38.6)

Notes:

(1) The estimates are the average of 1000 draws of the predicted wage with a random term. The 90% prediction interval over the 1000 draws is shown in parentheses.

(2) Single men's wages are estimated based on linear predicated wages. For predicted wages with a random term, we add a random disturbance drawn from the normal distribution $N(0, \hat{\sigma}_{\varepsilon}^{2})$ where $\hat{\sigma}_{\varepsilon}^{2} = \frac{RSS}{n-1}$ and RSS is the residual sum of squares from the linear regression over the sample of workers.

(3) All cases are for individuals 35 years old, living in Auckland, European and no other household income.

(4) Estimated wage rates are adjusted by Average Weekly Earnings to December quarter 2011.

Appendix B – Predicted wage distributions

Figure B.1 – Predicted log wage distribution for sole parents

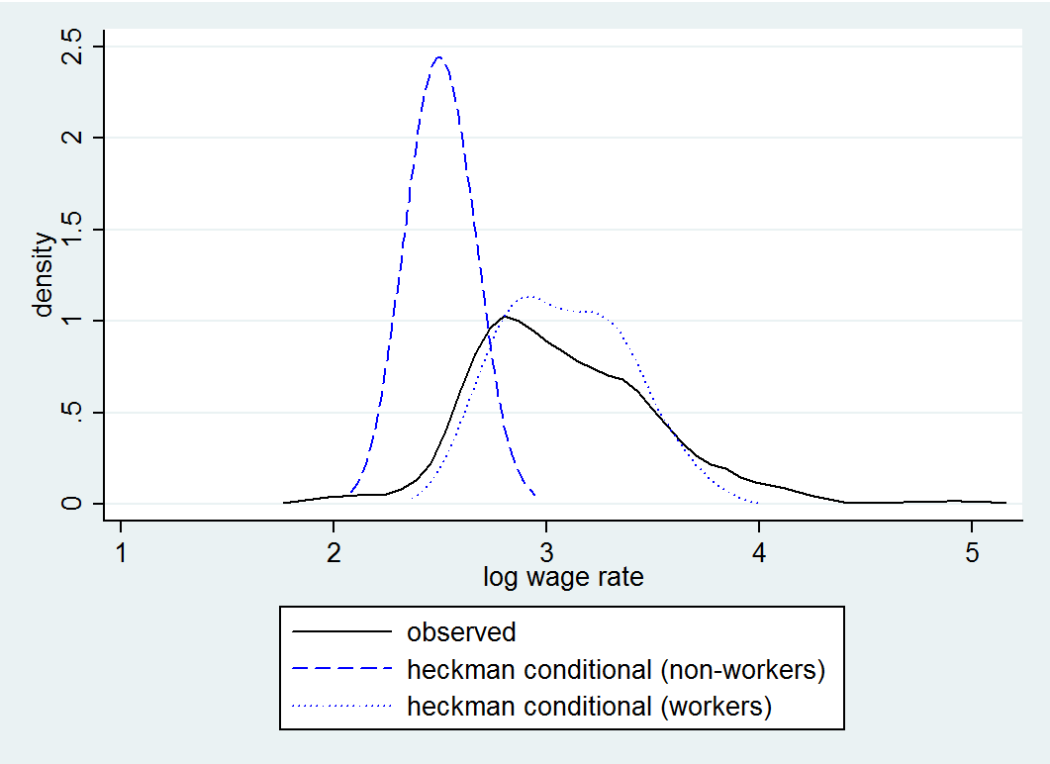


Figure B.2 – Predicted log wage distribution for sole parents: random term

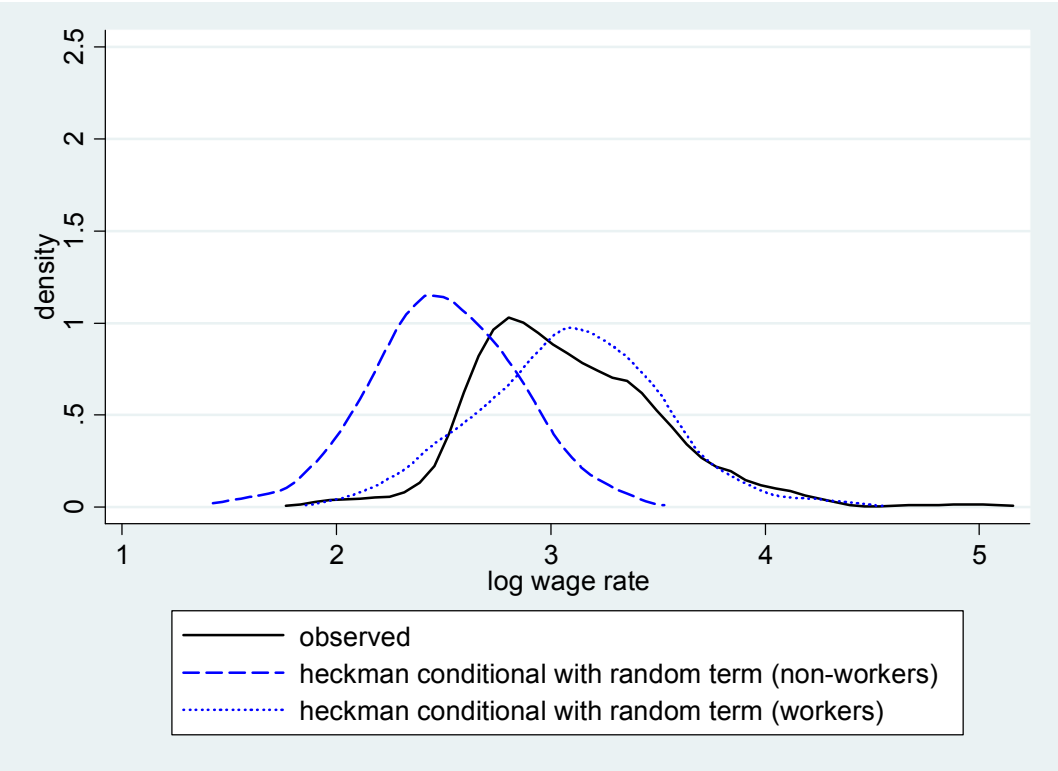


Figure B.3 – Predicted wage distribution for sole parents

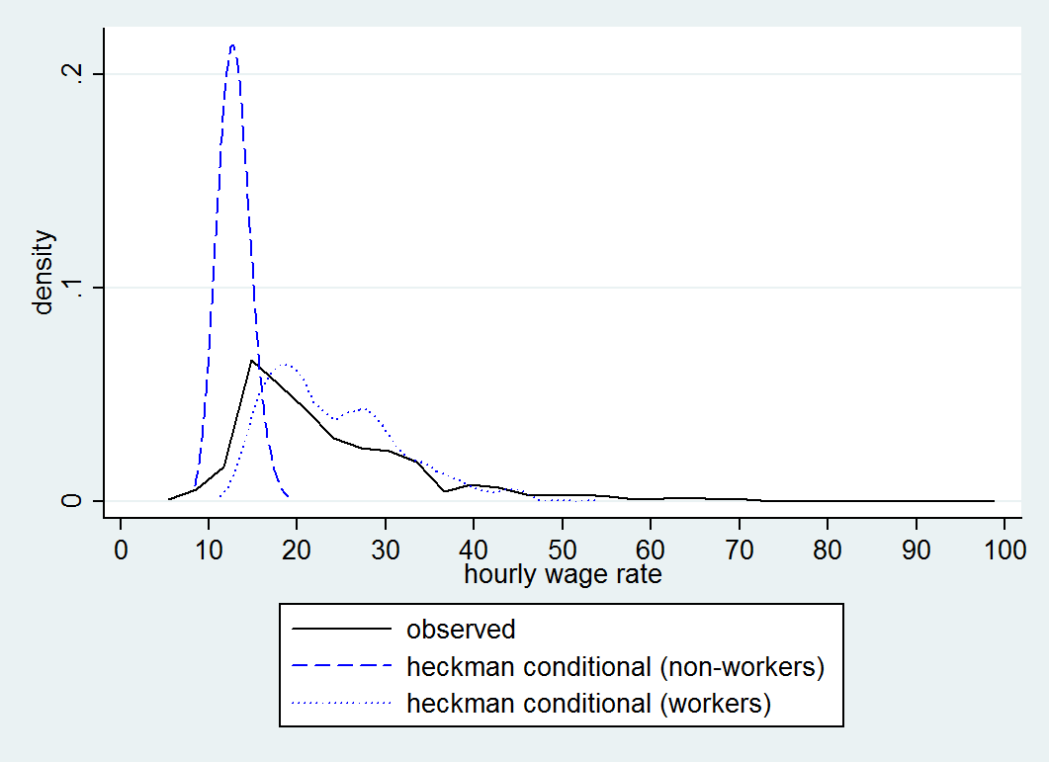


Figure B.4 – Predicted wage distribution for sole parents: random term