THE ACCURACY OF PREDICTED WAGES OF THE NON-EMPLOYED AND IMPLICATIONS FOR POLICY SIMULATIONS FROM STRUCTURAL LABOUR SUPPLY MODELS

Robert Breunig* and Joseph Mercante

Treasury Working Paper

2009 — 03

March 2009

*Corresponding author: Robert Breunig, Economics Program, Research School of Social Sciences, Building 9, Australian National University, Canberra, ACT 0200 Australia; E-mail: Robert.Breunig@.anu.edu.au. Robert Breunig would like to thank the Australian Treasury for their hospitality and support. We benefitted from comments by Deborah Cobb-Clark, Jennifer Foster, Paul Frijters, Xiaodong Gong, Guyonne Kalb, Anthony King, Paul Miller, Maryanne Mrakovcic and Marty Robinson. We are grateful to them. This paper uses confidentialised unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) survey. The HILDA Project was initiated and is funded by the Commonwealth Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (MIAESR). The findings and views reported in this paper, however, are those of the authors and should not be attributed to either FaHCSIA or the MIAESR or the Australian Treasury.

ABSTRACT

The main focus of this paper is on the accuracy of predicted wages for the nonemployed. We first examine whether the three groups of non-employed–the unemployed, the marginally attached, and the not in the labour force–should be modelled separately or together. We conclude that these are three distinct states and that they should not be pooled in modelling wages. We predict wages separately for the three non-employed groups using a range of two-state and four-state sample selection models. Using a panel data set from Australia, we test the accuracy of predicted wages for the non-employed by focusing on those individuals who subsequently enter employment. We find that conditional predictions, which incorporate the estimated sample selection correction, perform poorly for all groups, especially for the marginally attached and the not in the labour force. Unconditional predictions from the sample selection models perform better but never out-perform a simple linear regression. These results may have important implications for policy simulations from structural labour supply models.

JEL Classification Numbers: C52,J22,J30,J64,

Keywords: Labour supply; Selection bias; Multinomial logit; Wage predictions

1 Introduction

Labour supply models are often used to predict responses of individuals to changes in government tax and transfer systems. Of particular interest in many developed countries is the effect of such changes on individuals who are not currently working. Many government programs around the world, such as earned income tax credits and increased tax free income thresholds for low earners, are specifically designed to attract new workers into the work force and into employment.

An important aspect of the predictions from these labour supply models is the predicted wages which are generated for non-employed individuals. These predicted wages directly determine the additional (predicted) utility that non-employed individuals will get from working and thus the predicted changes in employment which will ensue from a policy change. For example, labour supply models will overstate (or understate) the employment benefits from tax cuts if wages of the non-working are systematically over-predicted (or under-predicted).

Australian examples of the use of predicted wages in structural labour supply modeling include Duncan and Harris (2002), Kalb (2002), Kalb and Lee (2008) and Breunig et al. (2008). Predicted wages, corrected for sample selection, are used in a wide variety of other applications beyond strucutral labour supply modeling. For example, Rammohan and Whelan (2005) generate predicted wages for modeling the choice of child care usage for working and non-working women.

The focus of this paper will be on two specific aspects of wage modelling for the non-working. First, we examine whether the unemployed, the marginally attached, and those not in the labour force should be treated identically or separately in modelling the probability of employment. We propose a new test for determining whether the non-employed should be categorized as one, two, or three groups. We find evidence that the unemployed, the marginally attached, and the not-in-the-labour force should be treated as three distinct groups for modelling purposes.

In the second part of the paper, we examine the wage predictions resulting from regressions which correct for selectivity bias using binomial and multinomial models of employment status. Specifically, we evaluate both conditional and unconditional wage predictions from these models. Using a panel of data from Australia, we compare predicted wages for the non-working to the wages they actually receive when they subsequently enter the labor market.

Overall, we find that wage predictions from wage equations which control for selection and which use information from the selection correction perform poorly. Selection correction terms are often poorly estimated and in small samples can be highly variable. For some groups that we consider, this results in very poor predictive performance. Including the estimated selection parameter in the wage equation leads to under-prediction of wages for the not in the labour force and marginally attached groups. For the unemployed, the results are more mixed, but it is clear that using the conditional (on selection) predictor sometimes produces very poor predictions. The main conclusion from the paper is that there is no compelling reason to use a conditional predictor for wages and that, in fact, there is very little gain in even using a selection correction model when generating predicted wages for the non-employed.

The paper is organized as follows. In section 2, we discuss wage models which control for selection into employment. In section 3 we discuss our data. In section 4, we discuss our strategy for testing whether the non-employed should be pooled or considered separately. In section 5, we examine the wage predictions from our models and compare them to realized wages for those who transition from not working to employment. We test which models generate the most accurate wage predictions. In section 6 we discuss our results and conclude.

2 Wage models with selection

The standard approach in the literature is that proposed by Heckman (1979) whereby wages w_i^* for all workers and non-workers depend upon a vector of observable human capital characteristics, $\mathbf{x}_i$ and some unobservable variables captured by u_i

$$ln(w_i^*) = \mathbf{x}_i'\boldsymbol{\beta} + u_i \tag{1}$$

The actual wage, w_i, is only observed if a latent variable $s_i^* > 0$ where

$$s_i^* = \mathbf{z}_i'\boldsymbol{\gamma} + v_i \tag{2}$$

$\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are vectors of parameters and equation (2) provides a model for the probability of employment. This latter equation captures the benefits of employment and thus

$\mathbf{z}_i$ must contain all of the variables in $\mathbf{x}_i$. If we think of this model as arising in the context of the Heckman (1974) reservation wage model, it should also contain variables which affect the reservation wage, which is (at least partially) determined by the costs of employment. Importantly, u_i and v_i are assumed to be jointly, normally distributed.

The two-step empirical approach is to estimate $\boldsymbol{\gamma}$ in (2) and use those to estimate

$$ln(w_i) = \mathbf{x}_i'\boldsymbol{\beta} + \rho\lambda\left(\mathbf{z}_i'\widehat{\boldsymbol{\gamma}}\right) + u_i \tag{3}$$

on the sample with observed wages. The inclusion of the inverse Mills ratio, λ, corrects for the fact that $E[v_i|s_i^* > 0] \neq 0$. In a reservation wage model, ρ captures two effects. The first effect is that unobservable characteristics which result in a higher wage will also result in a higher probability of employment. ρ will also capture the difference between the variance of wage offers and the covariance between wage offers and reservation wages. The first effect will be positive. The second effect will be negative if the covariance between reservation wages and wage offers, which one would expect to be positive, is greater than the variance of wage offers–see Ermisch and Wright (1994). Empirically, it is not rare for the latter effect to dominate and produce negative estimates of ρ.[1]

To predict wages from equation (3), one has several options. The unconditional predictor

$$E[ln(w_i)] = ln(\widehat{w}_i) = \mathbf{x}_i'\widehat{\boldsymbol{\beta}} \tag{4}$$

gives the best estimate of the wage for the case where we do not know whether or not the individual is working. If we know that the individual is working, we can condition on this information and use our model estimates to generate a conditional predicted wage for working individuals

$$E[ln(w_i)|s = 1] = ln(\widehat{w}_i^e) = \mathbf{x}_i'\widehat{\boldsymbol{\beta}} + \widehat{\rho}\frac{\phi\left(\mathbf{z}_i'\widehat{\boldsymbol{\gamma}}\right)}{\Phi\left(\mathbf{z}_i'\widehat{\boldsymbol{\gamma}}\right)}. \tag{5}$$

For those who are not employed, the conditional prediction of wages will be

$$E[ln(w_i)|s = 0] = ln(\widehat{w}_i^{ne}) = \mathbf{x}_i'\widehat{\boldsymbol{\beta}} + \widehat{\rho}\frac{-\phi\left(\mathbf{z}_i'\widehat{\boldsymbol{\gamma}}\right)}{1 - \Phi\left(\mathbf{z}_i'\widehat{\boldsymbol{\gamma}}\right)}. \tag{6}$$

[1] Dolton and Makepeace (1987) also discuss the difficulty of interpreting sample selection effects and point out that it is erroneous to argue that participants have lower earnings potential than non-participants when ρ is negative.

Note that in using equations (5) and (6) we are conditioning on unobservable human capital characteristics and on the relationship between the distributions of wage offers and reservation wages.[2]

If the model is correctly specified, the conditional predictor contains more information than the unconditional and Vella (1988) suggests its use in generating predicted wage gaps for black-white or male-female differences which condition on the work decision variables and the estimate of the parameter ρ. Use of the unconditional predictor provides only an estimate of the wage gap experienced by those who work. Schaffner (1998) points out that using the unconditional predictor is only valid under very restrictive conditions. In particular, if there are unobserved traits that matter for one group and not for the other, then wage gap estimates will be biased. Our focus will be on prediction for individuals rather than groups and the key assumption in using the conditional predictor is that the distribution of unobservables, in particular, the variances and covariances captured by ρ, are reasonably constant over time.

Puhani (2000) reviews some critiques of the Heckman selection approach. The approach does not provide an improvement in predictive power (for worker's wages) relative to ordinary least squares regression on the selected sample. It also suffers from potential collinearity problems when the variables in $\mathbf{z}_i'$ do not differ much from those in $\mathbf{x}_i'$. Lastly, the Heckman approach makes strong distributional assumptions which, when violated, may lead to poor model performance as has been validated in a number of monte carlo studies. These specification problems and the sensitivity of results to the strong model assumptions are generally found to be worse in small samples.[3]

In practice, one can estimate this model by pooling the unemployed, the marginally attached and the not in the labour force to form the category of non-workers or one can exclude one or more of these categories.[4] Flinn and Heckman (1983) find that, for young men, the unemployed and the not in the labour force are distinct groups and that the unemployment state facilitates job search in line with standard search theory models. Similar results are found by Tano (1991) for young people compared to older people,

[2]Hoffmann and Kassouf (2005) derive the marginal effects in a log earnings equation using the conditional predictor.

[3]Stolzenberg and Relles (1997) provide some intuition about specific mechanisms which can cause poor performance when using the Heckman selection approach.

[4]Most Australian studies treat the unemployed and the not in the labour force (including the marginally attached) as a combined group of non-workers. An exception is Ross (1986).

and Gonal (1992) for young women compared to young men. We will test whether the unemployed, the marginally attached and the not in the labour force are distinct groups and we will also check whether the distinction makes any difference in accurately predicting wages. These tests are described below.

If non-employment can best be described as a set of distinct categories, there may be predictive gains in modelling them as such. In that case, several methods have been suggested.[5] We begin with a multinomial model with J states. Each state $j = 1, \ldots, J$ has an associated utility which is described as

$$s_{ij}^* = \mathbf{z}_i' \boldsymbol{\gamma}_j + v_{ij}. \tag{7}$$

Without loss of generality, letting $j = 1$ be the employed state, wages are observed whenever

$$s_{i1}^* > \operatorname*{Max}_{j \neq 1} \{s_{ij}^*\}. \tag{8}$$

When the v_{ij} are independently and identically Gumbel distributed this produces the multinomial logit model (see McFadden (1973).)

The approach of Lee (1983) is to specify a bivariate distribution between u_i in equation (1) and ε_1, defined as

$$\varepsilon_1 = \max_{j \neq 1} \left(s_j^* - s_1^* \right) \tag{9}$$

with no restriction on the parametric form of the bivariate distribution beyond standard regularity conditions. Lee (1983) further assumes that the joint distribution of u and the inverse cumulative normal transformation of the cumulative distribution function of ε_1 do not depend upon the parameters of the distribution function of ε_1. In most applications, u_i is assumed to be normally distributed which implies a linearity restriction on the conditional distribution of u as discussed in Bourguignon et al. (2007, page 177).

Schmertmann (1994) shows that these assumptions imply very strong restrictions on the correlation between u and the v_j from equation (7). The correlations between the difference in unobservable determinants of the choice of alternative 1 against any other alternative and the unobservable determinants of wage must all have the same sign. If the unobservable determinants of utilities are identically distributed, as in the multinomial logit model, then these correlations must in fact be identical. Nonetheless, despite the restrictiveness of these assumptions, many empirical studies follow this route.

[5] In this paper, we do not consider nested models, where a sequence of choices are made.

Combining the approach of Lee (1983), with the multinomial logit model, and the normality assumption on the unobservables in equation (1) we estimate a wage equation, correcting for selection as

$$ln(w_i) = \mathbf{x}_i'\boldsymbol{\beta} - \sigma\rho\frac{\phi\left(\Phi^{-1}\left(F_1\left(\widehat{\gamma}_1, \ldots, \widehat{\gamma}_j\right)\right)\right)}{F_1\left(\widehat{\gamma}_1, \ldots, \widehat{\gamma}_j\right)} + u_i \tag{10}$$

where the γ_j are the estimated coefficients of the multinomial logit model and F_1 is the cumulative distribution function of the first alternative (employment). ϕ and Φ are the probability density function and cumulative distribution function, respectively, of the standard normal. Φ^{-1} is the inverse cumulative distribution function of the standard normal. w_i is only observed if workers are in the employed state. σ is the standard deviation of the unobservables from equation (1) and ρ is the correlation between those unobservables and the translation of v_{i1} from equation (9). We can not estimate ρ and σ separately, but the product of the two is estimated.

Once the parameters of equation (10) are estimated, one can use the estimate of $\boldsymbol{\beta}$ to predict wages using the unconditional predictor of equation (4). Alternately, one can create a conditional predictor for an individual's wage in state $j \neq 1$. The conditional predictor makes use of the extra information in $\widehat{\sigma\rho}$ and the estimates of F_1.

Another approach using the multinomial logit, proposed by Dubin and McFadden (1984), imposes a linearity assumption on the relationship between the error terms in the wage equation and the selection model. This gives rise to a wage equation, corrected for selection, as

$$ln(w_i) = \mathbf{x}_i'\boldsymbol{\beta} + \sigma\frac{\sqrt{6}}{\pi}\sum_{j=2}^{M} r_j\left(\frac{P_j ln(P_j)}{1 - P_j} - r_1 ln(P_1)\right) + u_i \tag{11}$$

r_j is the correlation between u_i in equation (1) and v_{ij} in equation (7) for the jth alternative.

If one assumes, as Dubin and McFadden (1984) do, that the correlations sum to zero across all states, then a restricted model may be estimated as

$$ln(w_i) = \mathbf{x}_i'\boldsymbol{\beta} + \sigma\frac{\sqrt{6}}{\pi}\sum_{j=2}^{M} r_j\left(\frac{P_j ln(P_j)}{1 - P_j} + ln(P_1)\right) + u_i \tag{12}$$

The linearity assumption proposed by Dubin and McFadden (1984) restricts the class of allowable distributions for u and imposes a specific form of linearity between u and

Gumbel distributions, see Bourguignon et al. (2007, page 179). This restriction does not allow for u to be normally distributed.

Bourguignon et al. (2007) propose an alternative restriction which allows normality of u. This restriction requires that the expected value of u conditional on v_1 through v_J be a linear function of the correlations between u and each v. This has the drawback of not providing a closed form solution for the conditional expectations of the v_1 through v_J, but the numerical computation is not particularly difficult.[6]

The wage, conditional on choosing to work, is

$$ln(w_i) = \mathbf{x}_i'\boldsymbol{\beta} + \sigma \left[r_1^* m(P_1) + \sum_{j=2}^{M} r_j^* m(P_j) \frac{P_j}{1-P_j} \right] + u_i \tag{13}$$

where the $m(P)$ are defined as

$$m(P_j) = \int \Phi^{-1} \left(z - ln(P_j) \right) g(z) dz \tag{14}$$

and the g are the probability density function of the v which are assumed to be identically distributed. r_j^* is the correlation between u and $\Phi^{-1}(v_j)$. For the predicted wages of non-working individuals, we can again use an unconditional or a conditional predictor.

We will use the four methods discussed above to predict wages for those who are not working and compare them to the actual observed wages that those same individuals earn once they enter the labour force. We do this using panel data, which we describe in the next section.

3 Data

The data are derived from the Household, Income and Labour Dynamics in Australia Survey (HILDA).[7] The HILDA Survey is a nationally representative annual panel survey of Australian households and we use the first five waves from 2001 to 2005. There are around 7,500 households and around 13,000 responding individuals in each wave. After removing multi-family households, same-sex couple households and couple households where partner information is unavailable, there are, in wave five, 3,954 married women and men, 695 lone parents, 1,108 single women and 989 single men.

[6] In implementing this method in section 5 below, we use the STATA code of Bourguignon et al. (2007) available at the link provided in their paper.

[7] See Watson and Wooden (2002) for more details.

We further restrict our sample to persons between 25 and 59 years of age, in order to exclude those facing decisions about full-time study or retirement. We drop the self-employed, workers in family businesses, full-time students and the retired. Also dropped are those receiving disability support pension, Department of Veteran's Affairs disability pension or sickness allowance. Finally persons who report working positive hours but state a zero wage are removed.[8] For couples, we drop the observation if either member satisfies one of these conditions. The analysis sample contains 1,492 married women and married men. In the final sample of 484 lone parents, the majority (88 percent) are women. Also there are 315 single women and 380 single men. The numbers are fairly similar for the earlier waves.

We discuss the definition of our key variables. Hours of labour supplied is defined as usual weekly hours of work in all jobs. The wage rate is defined as the person's gross weekly salary and wage income for all jobs divided by hours. For those not working, a wage of zero is assigned. Non-labour income is defined as the difference between gross income and salary and wage income over the financial year. Welfare income is income from pensions and benefits, family tax benefit, maternity allowance and childcare benefit.[9]

We categorize people into four employment states: employed (E), unemployed (U), marginally attached (M) and not in the labour force (NILF). A person is considered to be marginally attached to the labour force if they want to work and are actively looking for work but not available to start work in the reference week; or want to work and are not actively looking for work but are available to start work within four weeks. In Australian official statistics, as in most countries, the marginally attached are included in the NILF group. There is a growing literature across a range of countries (e.g., Gray et al. (2005) for Australia, Brandolini et al. (2006) for Europe, and Jones and Riddell (1999) and Jones and Riddell (2006) for Canada) showing that the three groups of non-

[8]Less than one per cent of the working sample reported zero wage.

[9]Where data for unearned income and salary and wage income were missing, we used the imputed values provided by HILDA. The imputation method for the first two waves is described in Watson (2004) and subsequent improvements based upon the method of Little and Su (1989) are discussed in Starick and Watson (2007). Following Frick and Grabka (2007), we included in our models dummy variables for each potentially imputed variable (wage, partner's wage, unearned income) which were set to one when we used an imputed value rather than the actual value. These imputation dummy variables made no difference to the results and we present the results without them.

employed behave quite differently in their propensity to transition to employment, with the marginally attached being less likely than the unemployed, but more likely than the NILF, to transition to employment. Appendix Table A1 provides details on the wave-by-wave sample sizes by labour force status.

Of particular interest in this study are the individuals who enter employment from one of the three non-employed categories. In our analysis sample, there are 581 cases (561 unique individuals) in the first five waves of HILDA where the individual is employed at time period $t+1$ and not employed at time t.

Table 1: Number of individuals in analysis sample entering employment by wave and by employment state in previous wave (full sample)

	Status in previous wave			
Employed in Wave	**Unemployed**	**Marginally Attached**	**NILF**	**Total**
2	58 (43.3%)	58 (23.4%)	47 (15.2%)	163 (23.6%)
3	53 (46.1%)	46 (23.0%)	40 (13.5%)	139 (22.8%)
4	42 (46.7%)	42 (24.1%)	47 (16.5%)	131 (23.9%)
5	46 (52.9%)	52 (32.9%)	50 (18.1%)	148 (28.4%)
Total	199 (46.7%)	198 (25.4%)	184 (15.8%)	581 (24.5%)

Data Source: Analysis sample from HILDA (see section three)

The percentages in Table 1 indicate the fraction of individuals from the particular employment state who transitioned to employment. For example, of the unemployed in wave one, 43.3% were employed in wave two. Table 2 provides the transitions by gender and single/partnered status. Throughout, we treat those in de-facto relationships as married. For non-partnered individuals, we separately consider lone parents. In the analysis of sections 4 and 5, we pool single males and females due to the small sample sizes in those groups.

Appendix Tables A2 and A3 provide population estimates from Australian Bureau of Statistics (2007) for monthly transitions to employment. On average across the six years, about 22 per cent of individuals who are unemployed transition to employment in a given month and about 6.7 per cent of those not in the labour force transition to employment. One would expect annual transitions to be higher, which is what we find. Comparison is rendered difficult as, in the official statistics, the marginally attached and NILF are combined and we are not able to separate out the two categories. Another problem is that the official statistics have not been subjected to the various sample exclusions that we have applied to the HILDA data.

Table 2: Number of individuals in analysis sample entering employment from non-employed state by household type

	Status in previous wave			
Subgroup	**Unemployed**	**Marginally Attached**	**NILF**	**Total**
Married males	67 (60.9%)	21 (32.8%)	10 (19.2%)	98 (43.4%)
Married females	49 (48.0%)	112 (26.7%)	145 (17.0%)	306 (22.3%)
Single males	35 (39.8%)	12 (27.9%)	3 (17.7%)	50 (33.8%)
Single females	17 (41.5%)	7 (31.8%)	5 (22.7%)	29 (34.1%)
Lone parents	31 (36.5%)	46 (19.9%)	21 (9.3%)	98 (18.1%)
Total	199 (46.7%)	198 (25.4%)	184 (15.8%)	581 (24.5%)

Data Source: Analysis sample from HILDA (see section three)

We are particularly interested in the wages of individuals who become employed after exiting the unemployed, marginally attached and not in the labour force categories. These are given in Table 3 by wave and Table 4 by gender/partnered/lone parent split. The wages in Tables 3 and 4 are not corrected for inflation.

Table 3: Mean (median) hourly wages of individuals in analysis sample who transition to employment

From Wave	To Wave	Unemployed	Marginally Attached	NILF	Employed
1	2	19.6 (17.2)	17.2 (15.2)	18.6 (17)	21.8 (19.6)
2	3	16.4 (15)	15.3 (14.6)	17.8 (15.6)	22.7 (20.2)
3	4	20.3 (16.3)	18.9 (16.7)	21.8 (20)	23.6 (21)
4	5	20.1 (17.4)	20.3 (16.7)	20.4 (18.3)	25.2 (22.2)
Total		19.0 (16.1)	17.9 (15.6)	19.7 (17.4)	23.3 (20.7)

Data Source: Analysis sample from HILDA (see section three)

We can test, using t-tests, whether mean wages in Table 3 are statistically different depending upon previous labour force status, without consideration of any individual characteristics. For those working, wages for the individuals who were employed in the immediately preceding wave (the last column of Table 3) are statistically larger (at the 10% level in all cases, at much lower levels for most cases) than wages for those who transition to employment from any of the other labour force states. For the most part, wage differences between those who were previously not in employment are not statistically different from one another. The exception is that for waves two to three and the pooled data, we find that the employed who were perviously NILF have statistically larger wages than the employed who were previously marginally attached.[10]

Table 4: Mean (median) hourly wages of individuals in analysis sample who transition to employment

Subgroup	**Unemployed**	**Marginally Attached**	**NILF**	**Employed**
Married males	21.1 (16.3)	18.4 (15.6)	18.7 (17)	26.0 (23)
Married females	18.3 (15.6)	18.3 (16.6)	20.3 (18)	21.0 (19)
Single males	19.6 (17.5)	20.1 (19.8)	15.6 (14.4)	22.7 (20)
Single females	17.3 (15.8)	21.5 (16.3)	16.9 (17.8)	21.5 (20)
Lone parents	15.9 (14.3)	15.7 (14.8)	17.8 (16.4)	20.7 (18.8)
Total	19.0 (16.1)	17.9 (15.6)	19.7 (17.4)	23.3 (20.7)

Data Source: Analysis sample from HILDA (see section three)

Turning to the outcomes classified by sex and marital status of Table 4, current mean wages for the previously employed are statistically greater than wages for the previously unemployed at the six per cent level or lower for all groups. For married males, married females, and lone parents, wages for the previously employed are statistically greater than wages for the previously marginally attached. The sample sizes for single males/females are small and it is difficult to make any statistical statement about these two groups of marginally attached. Wages for the previously employed are statistically greater than wages for the previously NILF for all groups except married females. Wages for the three groups of previously non-employed are not statistically different from one another for any of the sub-groups. The tests for equality of medians reveals the same patterns.

We draw several conclusions from the data on transitions into employment and wages for those who become employed. First, and in keeping with Flinn and Heckman (1983)

[10] If we conduct a non-parametric test of the equality of the medians, we find similar results. The median wages of the previously employed are significantly greater than those of the previously not employed for all three sub-groups.

and the subsequent literature that they inspired, we find that unemployed individuals have a higher probability of entering employment relative to those not in the labour force. We see this in both the monthly and the annual transitions. Secondly, it appears that our estimation sub-sample in HILDA has above-average propensity to become employed compared to population estimates of the Australian Bureau of Statistics. This is perhaps not surprising given the additional sample exclusions that we have made (full-time students and disabled) and the fact that we have annual, not monthly, transitions.

Married and single females have higher rates of movement from not in the labour force to employment relative to males. For married females this accords with our prior expectations. For married females and lone parents, average wages in employment when the previous state was not in the labour force are higher than wages when the previous state was either unemployment or marginal attachment, if we pool these two last categories. This result is significant at the 10% level in a one-sided test. This would be consistent with a model where married women and single parents who are caring for children at home have higher average labour productivity than unemployed women upon entering the workforce.

4 Should we treat all non-workers identically?

In this section we want to examine whether the three groups of non-workers, the unemployed, the marginally attached, and the not in the labour force, should be modeled separately or together. There is a growing literature which demonstrates that these three groups have very different propensities to become employed–see Jones and Riddell (1999) and Jones and Riddell (2006) for Canada, Brandolini et al. (2006) for Europe, and Gray et al. (2005) for Australia. We wish to address a different but related question: should the non-employed be considered as one, two or three separate groups when estimating a wage equation which corrects for sample selection?

We offer a new method to address this question, which is to specifically look at models of employment probability for these three groups in combination with the employed. To our knowledge, the classification tests that we propose below are new.[11] The

[11] Since writing this paper we have become aware of the paper of Ahn and Low (2007) who propose a similar approach to distinguishing between the unemployed and the not in the labour force. They do not separately consider the marginally attached.

advantage of these tests is that they directly address the question of which modelling approach of those discussed in section 2 above is appropriate–a binomial classification of the employed and non-employed and the Heckman model or a richer multinomial classification in conjunction with the Lee or McFadden methods.

Gray et al. (2005) have applied the tests of transition probabilities proposed by Jones and Riddell (1999) to Australia using a different data set which covers the period 1994 to 1997 and find that the marginally attached are distinct from both the unemployed and the not in the labour force. We applied these tests and the non-parametric tests of Brandolini et al. (2006) to our data and we also reject the hypothesis that the probabilities of transitioning into employment are identical for any of the groups of non-employed. We thus confirm that the conclusions of Gray et al. (2005) are also found for the 2001-2006 period using the HILDA data. Figure 1 shows the transitions to employment by wave for all individuals in our analysis sample who are non-employed at wave 1.

Turning to our proposed classification tests, we examine three different possibilities: that the unemployed (U) and the marginally attached (M) can be pooled; that the unemployed and the not in the labour force (NILF) can be pooled, and that the marginally attached and the not in the labour force can be pooled. If we find that two of these groups can be pooled, we can subsequently test whether that pooled group can be pooled with the third remaining category.

For each of the three pairings which we test, we propose five different classification tests. We outline these below using the test for pooling the unemployed and the marginally attached as an example. Our testing approaches are based upon estimation of binomial and multinomial choice models. We estimate three probit models

P1 Estimate probability of being employed using E, U and M.

P2 Estimate probability of being employed using E and M.

P3 Estimate probability of being employed using E and U.

If the model which determines non-employment is the same for the unemployed and the marginally attached, then P1, P2, and P3 should all (asymptotically) give similar answers. However, P2 and P3 should be inefficient relative to P1, since they only use a portion of the data. The basic principle underlying the Hausman (1978) test (comparison of two sets of coefficients, one of which is consistently estimated under the null and the other which is efficiently estimated under the null) therefore applies[12].

Thus, our first two tests are:

T1 Hausman test comparing coefficients from P1 to those of P2

T2 Hausman test comparing coefficients from P1 to those of P3

We can also compare estimates from a multinomial choice model to those from a binary choice model. For this comparison, we estimate two logistic models

L1 Binary logit for probability of being employed using E, U and M.

L2 Multinomial logit allowing U and M to be two distinct states

Again, the Hausman principle applies and we have two Hausman-type tests that can be produced from these estimates

T3 Hausman test comparing coefficients for unemployed from L1 and L2

T4 Hausman test comparing coefficients for marginally attached from L1 and L2

We can also use the multinomial logit estimates to conduct a Wald test to see if the coefficients for the unemployed and marginally attached states are equal.[13]

T5 F-test of equality of the coefficients for U and M from L2.

The probit and logit models are estimated using age, age squared, a dummy for poor English-speaking ability (self-assessed), a dummy variable for being in New South Wales, a dummy for living in a capital city, dummies for educational attainment, experience, experience squared, partner's wage, total unearned household income, number of resident children less than age 5, resident children aged 5-14, resident children older than 14, and non-resident children, a dummy variable if the individual owns their own home, a dummy if the individual is a public tenant, and dummies for imputed household income and imputed partner's wage.

Table A5 in the appendix contains a list of all the variables used in these regressions and their means and standard deviations from the fifth wave of the data.[14] We exclude

[12]Note that this is akin to the approach taken in Hausman and McFadden (1984)

[13]Another alternative would be the LR test of Cramer and Ridder (1991). In practice, this gives very similar results to test T5 and we do not report those results here.

[14]Detailed descriptive statistics for waves one through four are in Appendix B.

any variables which do not vary for the sub-sample of interest (e.g., we exclude *male* from the sub-sample of married males). We estimated all models with indicator variables if any of the wage or unearned income data were imputed–see footnote 9.

For each of our four sub-samples (married females, married males, lone parents, and singles[15]) we conduct tests T1 to T5 on each wave of data. We also conduct the tests on the data pooled across all five waves. For the pooled models, we conduct the Hausman tests in two different ways. We use the standard variance matrix of parameters uncorrected for the clustering which is created by the presence of multiple observations on the same individual in the pooled sample. We also conduct the Hausman tests using a variance matrix which is corrected for clustering using a standard outer-product correction. Neither are strictly correct, as the former does not account for the clustering and the latter is not strictly theoretically consistent with the Hausman test. Conclusions from the tests are consistent across both methods, however.[16]

The test results for married females and lone parents are summarized in tables 5 and 6.[17] We focus primarily on these two groups in what follows for two reasons. First, they are the two largest groups in terms of the size of the non-employed pool and also in terms of the numbers who transition to employment. Secondly, they are a frequent focus of government policy. Given current high employment in Australia, recent reforms to the tax and transfer system have been designed, at least in part, to induce married females and lone parents who are not in employment to enter the labour force and to enter employment (see Centrelink (2008)).

For the sub-sample of married females, we find consistent evidence across all waves that the unemployed, the not in the labour force, and the marginally attached are three distinct categories. Over 80 per cent of the wave-by-wave tests show significant differences and we find significant differences for all of the tests where we pool the data across waves. For lone parents, we only find significant differences in the wave-by-wave tests about 20 per cent of the time. However, when we pool across waves, we find strong evidence that the unemployed, the not in the labour force, and the marginally attached

[15]We pool single males and females due to small sample sizes.

[16]We only report the results using the variance-covariance matrix which is not corrected for clustering.

[17]Results for married males and singles are available in appendix B. Because of the small sample sizes, we generally find no differences for the individual waves. However, we find that the three states are distinct in the pooled tests for both married males and singles.

are three distinct categories. The results of the wave-by-wave tests can be explained by the small sample sizes once we split the non-employed into three separate states. These small sample sizes are not such a problem for married females.

Table 5: Married Women
Can we pool the unemployed, the marginally attached, and not in the labour force?
Table contains p-values for test of equality of labour force states

	Wave					
Test[a]	1	2	3	4	5	Pooled
Can we pool the not in the labour force and the unemployed?						
T1	.00**	.05*	.21	.01**	.00**	.00**
T2	.00**	.00**	.00**	.00**	.00**	.00**
T3	.04**	.05**	.24	.00*	.00**	.00**
T4	.00**	.00**	.00**	.00**	.00**	.00**
T5	.00**	.00**	.00**	.00**	.00**	.00**
Can we pool the marginally attached and the unemployed?						
T1	.01**	.08**	.66	.00**	.00**	.00**
T2	.00**	.00**	.00**	.00**	.00**	.00**
T3	.03**	.10*	.61	.00*	.00**	.00**
T4	.00**	.06*	.19	.01**	.00**	.00**
T5	.00**	.05*	.37	.01**	.00**	.00**
Can we pool the marginally attached and the not in the labour force ?						
T1	.00**	.00**	.00**	.00**	.00**	.00**
T2	.00**	.02**	.00**	.12	.16	.00**
T3	.00**	.01**	.00**	.00**	.03**	.00**
T4	.00**	.20	.01**	.17	.02**	.00**
T5	.00**	.04**	.00**	.03**	.02**	.00**
Can we pool the unemployed and a combined group of the not in the labour force and the marginally attached?						
T1	.01**	.07*	.36	.02**	.00**	.00**
T2	.00**	.00**	.00**	.00**	.00**	.00**
T3	.12	.15	.41	.15	.09*	.00**
T4	.00**	.00**	.01**	.00**	.00**	.00**
T5	.00**	.00**	.03**	.00**	.00**	.00**

Data Source: Analysis sample from HILDA (see section three)
Null hypothesis is that the two labour force states can be pooled. Table contains p-values.
[a] The five hypothesis tests, T1-T5 are described in detail in the text.
The last panel of the table compares the not in the labour force as traditionally defined, including the marginally attached, to the unemployed.

Table 6: Lone parents
Can we pool the unemployed, the marginally attached, and not in the labour force?
Table contains p-values for test of equality of labour force states

	Wave					
Test[a]	1	2	3	4	5	Pooled
	Can we pool the not in the labour force and the unemployed?					
T1	.04**	.00**	.05**	.03**	.14	.00**
T2	.00**	.21	.58	.11	.01**	.00**
T3	.14	.66	.89	.08*	.34	.00**
T4	.00**	.52	.48	.30	.13	.00**
T5	.02**	.54	.42	.23	.12	.00**
	Can we pool the marginally attached and the unemployed?					
T1	.06*	.00**	.44	.03**	.28	.01**
T2	.00**	.34	.28	.01**	.02**	.00**
T3	.33	.94	.02**	.12	.49	.00**
T4	.04**	.84	.50	.05*	.24	.00**
T5	.09*	.89	.49	.12	.22	.00**
	Can we pool the marginally attached and the not in the labour force ?					
T1	.01**	.01**	.55	.47	.54	.00**
T2	.07*	.01**	.01**	.06*	.86	.00**
T3	.19	.21	.76	.79	.98	.00**
T4	.19	.26	.32	.53	.98	.00**
T5	.13	.17	.66	.65	.98	.00**
	Can we pool the unemployed and a combined group of the not in the labour force and the marginally attached?					
T1	.06*	.00**	.00**	.04**	.12	.00**
T2	.00**	.37	.27	.02**	.01**	.00**
T3	.23	.85	.86	.20	.37	.00**
T4	.00**	.88	.23	.14	.06*	.00**
T5	.02**	.88	.21	.14	.07*	.00**

Data Source: Analysis sample from HILDA (see section three)
Null hypothesis is that the two labour force states can be pooled. Table contains p-values.
[a] The five hypothesis tests, T1-T5 are described in detail in the text.
The last panel of the table compares the not in the labour force as traditionally defined, including the marginally attached, to the unemployed.

In the last panel of both table 5 and table 6, we combine the marginally attached and the NILF as is done in the official statistics and test whether this combined group can be pooled with the unemployed. We can conclude from those tests that this combined group is also statistically significantly different from the unemployed. As the wage predictions which we discuss in section 5 are often estimated from models using ABS data which combine these two groups, we provide this test.

The conclusion we draw from these results is that it is a mistake to pool the unemployed, the marginally attached and the not in the labour force and treat them identically in modelling the probability of employment. This conclusion points the way to two possible modelling strategies for wage equations which correct for selection into employment. The first, is to model the employed with each of the non-employed groups separately. This would suggest separate estimation of three Heckman selection models for the three different groups. The problem with this strategy is that it is not clear which set of estimates one should use for understanding and predicting wages for the employed. A second modelling strategy which follows from these tests is to control for sample selection using the multinomial choice models discussed in section 2 above[18].

In this paper, our main focus is on those who are not in employment. We examine in the next section whether the results we have presented have any implications for predicted wages for *non-workers*. For all three groups of non-workers, we will examine the predicted wages from the different estimation strategies. We will then use the individuals who transition from non-work to employment to test which of these different estimation strategies provides the most accurate wage predictions for those non-workers who subsequently take up employment.

5 The accuracy of predicted wages using various modelling approaches

In the previous section, we concluded that the unemployed, the marginally attached and the not in the labour force appeared to be distinct groups when modelling the probability of employment. In this section, we consider whether these results have any relevance for the accuracy of predicted wages for these three groups.

Our basic approach will be as follows. We will estimate a model for wages in a particular cross-sectional wave, say wave t. We will then use the estimated model to predict a wage, $\widehat{w}_{it}$ for a non-employed individual. We then use an adjustment factor

[18] Instead of generating predicted wages from a selection model which are then plugged back into the labour supply model, another alternative is to jointly model labour supply and the wage equation, with four possible labour market states, and simultaneously estimate wages and labour supply. For a three-state example, see **?**

(a_t) to account for wage inflation between waves t and $t+1$ to generate a predicted wage for individual i at time $t+1$ as

$$\widehat{w}_{i,t+1} = \widehat{w}_{it}\left(1 + a_t\right) \tag{15}$$

In the results presented below, we used the average increase in wages in our sample data between wave t and $t+1$ for the adjustment factor. We also experimented with using the inflation rate of average weekly earnings from the Australian Bureau of Statistics, but this did not affect our conclusions.

We separately consider our two main sub-groups of interest: married women and lone parents.[19] We examine eleven separate models for predicting the wages for each sub-group. For each model, we include all of the variables from Table A5 in the appendix. We exclude from the wage equation the variables relating to unearned income, partner's wage, resident and non-resident children, and home ownership status.

- M1 Linear regression using only the employed
- M2 Heckman selection model using whole sample and conditional predictor of equation (6).
- M3 Heckman selection model using whole sample and unconditional predictor of equation (4).
- M4 Heckman selection model using only non-working population of interest (unemployed, marginally attached or not in the labour force) and conditional predictor of equation (6).
- M5 Heckman selection model using only non-working population of interest and unconditional predictor of equation (4).
- M6 Lee selection model of equation (10) and the conditional predictor of wages.
- M7 Lee selection model of equation (10) and the unconditional predictor of wages.
- M8 The original multinomial model of Dubin and McFadden, equation (12), and the conditional predictor of wages.
- M9 The original multinomial model of Dubin and McFadden, equation (12), and the unconditional predictor of wages.
- M10 The restricted multinomial model of Bourguignon, Fournier and Gurgand, equation (13), and the conditional predictor of wages.
- M11 The restricted multinomial model of Bourguignon, Fournier and Gurgand, equation (13), and the unconditional predictor of wages.

[19] Results for married men and singles are available from the authors.

For each of these we test whether the average predicted wage ($\widehat{w}_{i,t+1}$ above) is equal to the average realized wage for the three groups which transition into employment out of unemployment, marginal attachment or not in the labour force.

These results are summarized in Tables 9 through 11 for married females and in Tables 12 to 14 for lone parents. The rows of the table present the average predicted wages for the group in question. The p-value of the test of equality between the predicted log wage and the actual, observed log wage for those that transition into employment are given just below the average predicted wages.[20] We also pool our predictions across all waves in column 6. Column 7 presents the pooled results, dropping wave 1. For married women, we find oddly large wages for those in wave 2 who were unemployed in wave 1 (see table 9). There appears to be some variability in responses to wage and income questions which settles down in subsequent waves as respondents become more adept at accurately completing the questionnaire. We dropped the wave 1 to wave 2 changes to see if our results were sensitive to any potential problem. In our discussion, we will focus primarily on the pooled results rather than the wave-by-wave results. For the latter, sample sizes are sometimes fairly small and this introduces variability into the results.

5.1 Discussion of results

We draw several conclusions from the results. The first conclusion is that the unconditional wage prediction from all of the models across all of the sub-groups is never statistically different to the wage prediction that one would make based upon a linear regression model estimated only on the sub-population of working individuals.

The second unambiguous conclusion from the results is that the conditional predictor which uses the estimated sample selection parameter in the prediction is highly variable. This is particularly true for the multinomial models where some of the conditional wage predictions are unrealistic. It is also true for the Heckman correction model. Looking at the pooled results in the row labeled M4 in Table 9, for example, we see that average

[20] For ease of reading, we present the wages in levels. We have used a consistent predictor of the wage level based upon the estimates of the log wage model without imposing any parametric assumptions. As the model is estimated in log wage, we present the p-values of the test which compares predicted to actual log wage. We do this so that our tests are not influenced by the noise generated in estimating the scaling factor which we use to inflate $\exp(ln(wage))$ to wage level.

predicted wages are nearly twice the average actual wage shown in the first row of the table. This problem arises in part because the sample selection term is often estimated with very low precision. The estimates of the sample selection term are also unstable–switching between negative and positive for different waves of data using the same population.[21]

The third conclusion is that there is no obvious gain from using a multinomial model relative to a simple Heckman correction model. The conditional predictors from those models, as discussed above, are highly unstable. The unconditional predictors do not vary much from the unconditional predictor from the Heckman model nor from linear predictor from a regression on the selected sample.

Our fourth conclusion is that a simple linear predictor from a regression on the selected sample or the unconditional predictor from the sample selection model very often outperforms the conditional predictor which uses information from the sample selection correction. This is certainly the case for married women who move from not in the labour force to employment (M1, M3 and M5 in Table 10), for married women who move from marginal attachment to employment (M3 and M5 in Table 11), and lone parents moving from either unemployment or marginal attachment (Tables 12 and 14) to employment. For lone parents who transition from not in the labour force to employment, all of the techniques produce reasonable predictions.

For married women who move from unemployment to employment, the results are more mixed. The unconditional predictor works better (M3 of Table 9) across all waves, but if we consider only the last four waves, then the conditional predictor works better. Given the extreme observation for average wages for those who move from unemployment in wave 1 to employment in wave 2, we might prefer the conditional predictor for this group. But if we estimate the model only on the employed and unemployed (dropping the not in the labour force and the marginally attached), then the conditional predictor performs very poorly. This is probably due to the small sample size, but it is somewhat disturbing that the conditional predictor performs so differently in rows M2 and M4 of Table 9.

[21]For the Heckman selection models, Table A4 in the appendix provides a summary of the sign and significance of the estimated sample selection correction parameter.

Table 9: Predicted wages
Married women who transition from Unemployed (U) to employed

	From wave: **To wave:**	1 2	2 3	3 4	4 5	Pooled All waves	Pooled W2-W5 only
	Observed in data	24.97	16.71	15.68	16.89	18.35	16.39
	Predictions from different models						
M1	**Linear regression**	20.88 (0.60)	18.50 (0.48)	20.00* (0.09)	20.99 (0.16)	20.43* (0.08)	20.28** (0.02)
Heckman selection model using whole sample							
M2	**Conditional predictor**	14.42** (0.01)	15.68 (0.18)	18.44 (0.36)	19.57 (0.35)	16.67* (0.08)	17.75 (0.75)
M3	**Unconditional predictor**	20.14 (0.25)	18.16 (0.90)	19.82 (0.15)	20.77 (0.21)	19.91 (0.42)	19.92* (0.07)
Heckman selection model using employed and unemployed							
M4	**Conditional predictor**	35.47** (0.01)	27.53*** (0.00)	15.68 (0.62)	23.50** (0.04)	39.48*** (0.00)	38.87*** (0.00)
M5	**Unconditional predictor**	21.08 (0.70)	18.67 (0.37)	19.89 (0.11)	21.06 (0.15)	20.72** (0.03)	20.61*** (0.00)
Lee selection model							
M6	**Conditional predictor**	14.78** (0.02)	15.82 (0.21)	18.65 (0.30)	19.42 (0.38)	16.78* (0.10)	17.78 (0.74)
M7	**Unconditional predictor**	20.19 (0.26)	18.18 (0.87)	19.85 (0.14)	20.75 (0.21)	19.93 (0.40)	19.93* (0.07)
Original Dubin-McFadden model							
M8	**Conditional predictor**	365.57*** (0.00)	126.45*** (0.00)	25.88*** (0.00)	29.00*** (0.00)	92.67*** (0.00)	28.72*** (0.00)
M9	**Unconditional predictor**	20.48 (0.34)	18.91 (0.52)	19.97 (0.11)	21.01 (0.19)	20.34 (0.20)	20.45** (0.02)
Restricted Dubin-McFadden model							
M10	**Conditional predictor**	314.22*** (0.00)	494.63*** (0.00)	17.41 (0.70)	12.36* (0.05)	204.86*** (0.00)	92.80*** (0.00)
M11	**Unconditional predictor**	20.55 (0.40)	19.40 (0.16)	19.86 (0.16)	20.68 (0.29)	20.56** (0.05)	20.76*** (0.00)

Data Source: Analysis sample from HILDA (see section three)
Table entries are observed and predicted wages ($\widehat{w}_{i,t+1}$). We estimate a model in $ln(wage)$ but use a consistent predictor of the wage level from the $ln(wage)$ model.
Numbers in parentheses are p-values for tests of equality between average predicted log wage, $\ln(\widehat{w}_{i,t+1})$ and observed log wage at time $t+1$.
*** indicates significant difference between observed and predicted $ln(wage)$ at 1% level. ** and * indicate significance at the 5 and 10 % levels, respectively.

Table 10: Predicted wages
Married women who transition from Not in the labour force (N) to employed

	From wave: **To wave:**	1 2	2 3	3 4	4 5	Pooled All waves	Pooled W2-W5 only
	Observed in data	18.79	17.89	22.80	20.77	20.14	20.63
	Predictions from different models						
M1	**Linear regression**	18.46 (0.64)	18.70 (0.35)	19.64 (0.18)	21.67 (0.24)	19.80 (0.74)	20.31 (0.56)
Heckman selection model using whole sample							
M2	**Conditional predictor**	14.13*** (0.00)	16.19 (0.15)	18.43** (0.02)	20.43 (0.64)	16.79*** (0.00)	18.16** (0.02)
M3	**Unconditional predictor**	17.77** (0.04)	18.37 (0.87)	19.44* (0.09)	21.52 (0.33)	19.36 (0.20)	20.01 (0.77)
Heckman selection model using employed and not in the labour force							
M4	**Conditional predictor**	13.57*** (0.00)	15.40* (0.05)	18.42** (0.02)	19.35 (0.90)	15.88*** (0.00)	17.16*** (0.00)
M5	**Unconditional predictor**	17.96 (0.12)	18.38 (0.80)	19.50 (0.11)	21.43 (0.37)	19.38 (0.27)	19.98 (0.75)
Lee selection model							
M6	**Conditional predictor**	14.27*** (0.00)	16.36 (0.21)	18.57** (0.02)	20.29 (0.70)	16.87*** (0.00)	18.18** (0.03)
M7	**Unconditional predictor**	17.81** (0.05)	18.39 (0.83)	19.47* (0.09)	21.51 (0.34)	19.37 (0.22)	20.01 (0.77)
Original Dubin-McFadden model							
M8	**Conditional predictor**	15.96*** (0.00)	8.22*** (0.00)	16.77*** (0.00)	10.70*** (0.00)	10.27*** (0.00)	8.74*** (0.00)
M9	**Unconditional predictor**	17.85* (0.07)	18.73 (0.54)	19.54 (0.12)	21.49 (0.35)	19.62 (0.49)	20.31 (0.84)
Restricted Dubin-McFadden model							
M10	**Conditional predictor**	16.37*** (0.00)	9.35*** (0.00)	17.54*** (0.00)	10.57*** (0.00)	11.96*** (0.00)	10.40*** (0.00)
M11	**Unconditional predictor**	17.92 (0.14)	19.03 (0.10)	19.46* (0.07)	21.31 (0.56)	19.80 (0.58)	20.56 (0.15)

Data Source: Analysis sample from HILDA (see section three)
See footnotes to Table 9.

Table 11: Predicted wages
Married women who transition from Marginally attached (M) to employed

	From wave: / To wave:	1 / 2	2 / 3	3 / 4	4 / 5	Pooled All waves	Pooled W2-W5 only
	Observed in data	18.71	16.48	19.58	18.48	18.40	18.25
	Predictions from different models						
M1	**Linear regression**	18.51 (0.82)	18.82* (0.09)	18.57 (0.55)	21.59* (0.09)	19.52 (0.16)	20.13* (0.06)
Heckman selection model using whole sample							
M2	**Conditional predictor**	14.10*** (0.00)	16.21 (0.49)	17.30* (0.08)	20.30 (0.79)	16.45*** (0.00)	17.91 (0.23)
M3	**Unconditional predictor**	17.68* (0.06)	18.46 (0.31)	18.38 (0.34)	21.49 (0.14)	19.08 (0.79)	19.85 (0.32)
Heckman selection model using employed and marginally attached							
M4	**Conditional predictor**	10.90*** (0.00)	14.85 (0.13)	15.84*** (0.01)	25.54*** (0.00)	14.19*** (0.00)	17.56 (0.21)
M5	**Unconditional predictor**	17.64* (0.08)	18.58 (0.19)	18.34 (0.35)	21.70* (0.06)	19.15 (0.83)	19.99 (0.13)
Lee selection model							
M6	**Conditional predictor**	14.28*** (0.00)	16.35 (0.59)	17.45 (0.11)	20.16 (0.44)	16.53*** (0.00)	17.92 (0.24)
M7	**Unconditional predictor**	17.74* (0.07)	18.49 (0.28)	18.40 (0.36)	21.48 (0.14)	19.10 (0.83)	19.86 (0.31)
Original Dubin-McFadden model							
M8	**Conditional predictor**	4.51*** (0.00)	45.46*** (0.00)	20.98 (0.36)	113.75*** (0.00)	30.19*** (0.00)	121.80*** (0.00)
M9	**Unconditional predictor**	17.83 (0.11)	18.88 (0.16)	18.52 (0.45)	21.73 (0.10)	19.39 (0.67)	20.24 (0.11)
Restricted Dubin-McFadden model							
M10	**Conditional predictor**	7.62*** (0.00)	53.25*** (0.00)	17.50* (0.09)	91.96*** (0.00)	30.87*** (0.00)	99.00*** (0.00)
M11	**Unconditional predictor**	17.92 (0.20)	19.23** (0.03)	18.40 (0.29)	21.62 (0.22)	19.57* (0.09)	20.46*** (0.00)

Data Source: Analysis sample from HILDA (see section three)
See footnotes to Table 9.

Table 12: Predicted wages
Lone parents who transition from Unemployed (U) to employed

	From wave: **To wave:**	1 → 2	2 → 3	3 → 4	4 → 5	Pooled All waves	Pooled W2-W5 only
	Observed in data	14.67	11.00	16.71	20.67	16.14	16.52
	Predictions from different models						
M1	**Linear regression**	15.89 (0.72)	18.48^{*} (0.09)	19.65 (0.31)	17.83 (0.70)	17.87 (0.16)	18.50 (0.17)
Heckman selection model using whole sample							
M2	**Conditional predictor**	18.48^{*} (0.08)	13.63 (0.51)	20.56 (0.18)	16.00 (0.30)	19.03^{**} (0.04)	19.28^{*} (0.08)
M3	**Unconditional predictor**	16.53 (0.24)	17.38 (0.17)	19.84 (0.25)	16.92 (0.43)	18.17^{*} (0.09)	18.70 (0.12)
Heckman selection model using employed and unemployed							
M4	**Conditional predictor**	27.07^{***} (0.01)	12.39 (0.59)	25.59^{**} (0.02)	23.70 (0.26)	24.09^{***} (0.00)	26.18^{***} (0.00)
M5	**Unconditional predictor**	16.46 (0.29)	18.03 (0.11)	20.10 (0.21)	19.02 (0.96)	18.30^{*} (0.07)	19.02^{*} (0.08)
Lee selection model							
M6	**Conditional predictor**	18.68^{*} (0.07)	11.96 (0.46)	21.22 (0.12)	16.04 (0.30)	18.86^{**} (0.05)	19.28^{*} (0.08)
M7	**Unconditional predictor**	16.57 (0.23)	17.45 (0.17)	19.99 (0.21)	16.93 (0.43)	18.12^{*} (0.10)	18.70 (0.12)
Original Dubin-McFadden model							
M8	**Conditional predictor**	67.75^{***} (0.00)	6.07 (0.19)	71.04^{***} (0.00)	49.21^{***} (0.01)	29.67^{***} (0.00)	36.92^{***} (0.00)
M9	**Unconditional predictor**	16.77 (0.16)	17.57 (0.14)	20.42 (0.18)	18.89 (0.91)	18.69^{**} (0.04)	19.27^{*} (0.07)
Restricted Dubin-McFadden model							
M10	**Conditional predictor**	112.02^{***} (0.00)	4.12^{**} (0.04)	83.10^{***} (0.00)	71.73^{***} (0.00)	47.48^{***} (0.00)	60.89^{***} (0.00)
M11	**Unconditional predictor**	17.00^{*} (0.06)	17.08 (0.19)	20.86^{*} (0.08)	20.06 (0.64)	19.03^{**} (0.01)	19.74^{**} (0.03)

Data Source: Analysis sample from HILDA (see section three)
See footnotes to Table 9.

Table 13: Predicted wages
Lone parents who transition from Not in the labour force (N) to employed

	From wave: **To wave:**	1 2	2 3	3 4	4 5	Pooled All waves	Pooled W2-W5 only
	Observed in data	16.33	17.50	17.00	18.89	17.94	18.27
Predictions from different models							
M1	**Linear regression**	18.61 (0.98)	19.71 (0.51)	20.81 (0.55)	19.12 (0.28)	18.81 (0.71)	19.22 (0.84)
Heckman selection model using whole sample							
M2	**Conditional predictor**	22.09 (0.49)	14.88* (0.07)	21.64 (0.48)	17.18*** (0.01)	20.04 (0.28)	20.02 (0.46)
M3	**Unconditional predictor**	19.64 (0.61)	19.04 (0.80)	20.78 (0.54)	18.53** (0.05)	19.03 (0.84)	19.35 (0.89)
Heckman selection model using employed and not in the labour force							
M4	**Conditional predictor**	20.39 (0.74)	14.20* (0.09)	21.90 (0.47)	16.10*** (0.00)	19.01 (0.88)	18.92 (0.61)
M5	**Unconditional predictor**	18.93 (0.89)	19.34 (0.83)	20.74 (0.55)	18.63* (0.07)	18.83 (0.74)	19.19 (0.80)
Lee selection model							
M6	**Conditional predictor**	22.09 (0.47)	15.09* (0.08)	22.34 (0.45)	17.19*** (0.01)	19.86 (0.38)	20.03 (0.45)
M7	**Unconditional predictor**	19.67 (0.60)	19.13 (0.86)	20.76 (0.52)	18.54* (0.05)	19.00 (0.90)	19.35 (0.89)
Original Dubin-McFadden model							
M8	**Conditional predictor**	8.65 (0.11)	20.34 (0.53)	23.61 (0.30)	12.13*** (0.00)	8.33*** (0.00)	8.11*** (0.00)
M9	**Unconditional predictor**	19.69 (0.45)	19.67 (0.75)	20.22 (0.55)	19.76 (0.56)	19.31 (0.43)	19.58 (0.57)
Restricted Dubin-McFadden model							
M10	**Conditional predictor**	13.59 (0.58)	17.95 (0.23)	26.50 (0.32)	12.88*** (0.00)	11.57*** (0.00)	10.44*** (0.00)
M11	**Unconditional predictor**	20.18 (0.26)	19.27 (0.78)	20.13 (0.46)	20.51 (0.43)	19.58* (0.08)	19.88 (0.13)

Data Source: Analysis sample from HILDA (see section three)
See footnotes to Table 9.

Table 14: Predicted wages
Lone parents who transition from Marginally attached (M) to employed

	From wave: To wave:	1 2	2 3	3 4	4 5	Pooled All waves	Pooled W2-W5 only
	Observed in data	14.18	13.17	15.75	20.91	15.95	16.58
	Predictions from different models						
M1	**Linear regression**	16.14 (0.30)	18.31 (0.11)	19.09 (0.15)	20.70 (0.91)	18.35* (0.06)	19.20* (0.10)
Heckman selection model using whole sample							
M2	**Conditional predictor**	18.97*** (0.00)	12.82 (0.81)	19.87* (0.07)	18.54 (0.30)	19.59*** (0.01)	20.02** (0.03)
M3	**Unconditional predictor**	17.07** (0.03)	16.62 (0.45)	19.24 (0.12)	20.03 (0.59)	18.66** (0.02)	19.39* (0.06)
Heckman selection model using employed and marginally attached							
M4	**Conditional predictor**	21.61*** (0.00)	13.36 (0.89)	18.14 (0.31)	17.59 (0.21)	20.88*** (0.00)	21.09*** (0.01)
M5	**Unconditional predictor**	17.41** (0.02)	17.40 (0.25)	18.95 (0.17)	20.06 (0.64)	18.75** (0.02)	19.47* (0.06)
Lee selection model							
M6	**Conditional predictor**	19.15*** (0.00)	13.09 (0.90)	20.52** (0.04)	18.56 (0.31)	19.41*** (0.01)	20.03** (0.03)
M7	**Unconditional predictor**	17.10** (0.03)	16.73 (0.42)	19.35* (0.10)	20.03 (0.59)	18.61** (0.03)	19.40* (0.06)
Original Dubin-McFadden model							
M8	**Conditional predictor**	20.03** (0.01)	356.54* (0.09)	7.83*** (0.00)	16.28** (0.03)	39.60*** (0.00)	39.88*** (0.00)
M9	**Unconditional predictor**	17.86** (0.01)	17.21 (0.29)	19.32 (0.12)	21.44 (0.96)	19.18*** (0.01)	19.90** (0.03)
Restricted Dubin-McFadden model							
M10	**Conditional predictor**	24.51*** (0.00)	13.92 (0.91)	10.44*** (0.00)	31.82*** (0.00)	39.31*** (0.00)	42.37*** (0.00)
M11	**Unconditional predictor**	18.45*** (0.00)	17.03 (0.43)	19.65* (0.06)	22.18 (0.55)	19.52*** (0.00)	20.30*** (0.01)

Data Source: Analysis sample from HILDA (see section three)
See footnotes to Table 9.

The sample sizes for lone parents are smaller than for married females and thus the results may be less reliable. Across Tables 12 to 14 and all models we estimated, the linear predictor from the simple regression on the selected sample never performs worse than either the conditional or unconditional predictor from the sample selection models. The results for lone parents confirm the four main conclusions enumerated above.

One might worry that those non-employed individuals at period t who become employed at period $t+1$ are not a random sample from the group of non-employed, but are themselves a selected sample with unobservable characteristics better than the average non-employed person. In that case, our tests may be interpreted as a test for the best predictor of wages conditional on actually taking up employment in subsequent periods. For some types of policy simulations, this may be the relevant predicted wage.

It is very difficult to get a good estimate of the unobservable characteristics for those who never take up employment. For those who move from non-employment to employment, we can estimate the unobservable effects on wages through the residual from the wage regression at time $t+1$.[22] If we take the residuals from a wage regression estimated on the entire pooled sample of individuals who are employed and then run a regression on a set of dummy variables which indicate the previous employment status (one wave prior), we find significantly negative effects of having been either unemployed or marginally attached in the previous period.[23] The unobservables for the previously not in the labour force are less than those of the previously employed, on average, but the difference is not significant.

We find this result reassuring in regards to the amount of selection that might be present in our sample which moves from non-employment to employment. We expect, a priori, that the unemployed and the marginally attached might have poorer unobservable labour market characteristics than the employed and this is in fact what we find.

Taking the conclusions from sections 4 and 5, we plan to explore in future work the consequences of the use of different prediction techniques on the policy conclusions derived from a structural labour supply model. We have conducted some preliminary investigation using the Melbourne Institute Tax and Transfer Simulator (MITTS) model. However, our ability to fully implement our conclusions is limited since the model is based upon Australian Bureau of Statistics data which does not distinguish between the marginally attached and the not in the labour force. MITTS uses conditional wage estimates for all non-workers grouped together. Initial results show little change in labour supply estimates from the model when using separate wage imputation for the unemployed (based upon the conditional predictor) and for the combined group of marginally

[22] This will be independent of the estimate of the correlation between utility of employment and wages estimated at time t.

[23] We control for the clustering induced by the pooling of individuals across time.

attached and not in the labour force (using the unconditional predictor). Results using a model based upon a richer data source, such as HILDA, may prove to be different.

6 Discussion and Conclusions

In a model of the probability of employment, we find that the unemployed, the marginally attached and the not-in-the-labour force appear to be three distinct groups. This result is consistent across several different types of models and different specification tests. The implication is that these three groups should not be pooled together into one 'non-employed' group in a joint model of wages and employment. Our conclusion is based upon specification tests of cross-sectional models which classify individuals into one or another category. Looking at transitions to employment, Gray et al. (2005) are led to similar conclusions for Australia using data covering the period 1994 to 1997. Applying similar tests to the transitions in our data, we come to the same conclusion for the 2001 to 2005 period.

Building upon these results, we examine the wage predictions from a variety of models beginning with a simple linear regression model which has no controls for selection into employment to more complicated models which allow for multiple non-employment states. We find that the linear predictor from a regression on the selected sample of workers almost always out-performs more complicated prediction strategies. Interestingly, this is the same conclusion that is reached by Duan et al. (1983) for the problem of predicting the dependent variable for the *selected* sample. Our paper is the first to examine the question of predictive power for the *non-selected* sample.

The linear predictor doesn't always provide unbiased estimates of future wages, but it is less prone to very large errors than conditional prediction based upon a sample selection model. This is primarily driven by the instability and imprecision of the estimated coefficient on the sample selection term in the second stage of the two-step modelling procedure. More complicated multinomial models appear to suffer from these problems to a greater degree than the binary Heckman selection model.

A caveat to these general conclusions is that for married women who are unemployed, we do find some evidence that including the sample selection correction in wage predictions provides some improvement to the linear (unconditional) predictor. This result is somewhat sensitive to the sample period chosen and to sample size.

For married women in the not in the labour force and marginally attached categories, the selection model does not seem to provide information about the average effect of unobservables on wages through the correlation with the selection equation. One possible explanation is that the decision to move from one of these non-employed states to employment is accompanied by a change in the relationship between the reservation wage and the distribution of wage offers as discussed in section 2 above.

For unemployed, married women, however, there does seem to be information in the selection model regarding unobservables. This is consistent with a model that views unemployment as the state in which individuals better understand their reservation wage and truly are ready to take up employment if the right offer comes along.

Some labour supply models use predicted wages in policy simulations to consider the likely employment outcomes from changes to the tax and transfer system. Our paper provides several lessons for individuals who are estimating such models using survey data. The first is that for the marginally attached and the not in the labour force, prediction using a simple linear regression on the selected sample seems to out-perform any other potential method. The second conclusion is that a one-size-fits-all approach to predicting wages for those who are not working may be inappropriate. It may be appropriate to use conditional prediction for the unemployed, where there does seem to be useful information about unobservable influences on wage obtained through the correlation between the participation decision and wages.

References

Ahn, S. C. and Low, S. (2007). A parametric test for the distinction between unemployed and out of the labour force statuses. *Seoul Journal of Economics*, 20(1):59–87.

Australian Bureau of Statistics (2007). Labour Force, Australia. Detailed Electronic Delivery, March 2007–Gross Flows (ST GM1), Australian Bureau of Statistics. Datacube 6291.0.55.001.

Bourguignon, F., Fournier, M., and Gurgand, M. (2007). Selection bias corrections based on the multinomial logit model: Monte carlo comparisons. *Journal of Economic Surveys*, 21(1):174–205.

Brandolini, A., Cipollone, P., and Viviano, E. (2006). Does the ILO definition capture all unemployment? *Journal of the European Economic Association*, 4(1):153–179.

Breunig, R., Cobb-Clark, D., and Gong, X. (2008). Improving the modeling of couples' labour supply. *Economic Record*, 84(267):466–485.

Centrelink (2008). Welfare to work. Centrelink Publication, Commonwealth of Australia 2008. Available at: `http://www.centrelink.gov.au/internet/internet.nsf/services/welfare_work.htm`.

Cramer, J. and Ridder, G. (1991). Pooling states in the multinomial logit. *Journal of Econometrics*, 47:267–272.

Dolton, P. J. and Makepeace, G. H. (1987). Interpreting sample selection effects. *Economics Letters*, 24:373–379.

Duan, N., Manning, W. G., Morris, C. N., and Newhouse, J. P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics*, 1(2):115–126.

Dubin, J. A. and McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, 52(2):345–362.

Duncan, A. and Harris, M. (2002). Simulating the behavioural effects of welfare reforms among sole parents in Australia. *The Economc Record*, 78:264–276.

Ermisch, J. F. and Wright, R. E. (1994). Interpretation of negative sample selection effects in wage offer equations. *Applied Economics Letters*, 1:187–189.

Flinn, C. J. and Heckman, J. J. (1983). Are unemployment and out of the labour force behaviorally distinct labor force states? *Journal of Labor Economics*, 1(1):28–42.

Frick, J. R. and Grabka, M. M. (2007). Item-non-response and imputation of labor income in panel surveys from a cross-national comparison. IZA Discussion paper number 3043.

Gonal, F. (1992). New evidence on whether unemployment and out of the labor force are distinct states. *Journal of Human Resources*, 27(2):329–361.

Gray, M., Heath, A., and Hunter, B. (2005). The labour force dynamics of the marginally attached. *Australian Economic Papers*, 44(1):1–14.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, 46:1251–1271.

Hausman, J. A. and McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*, 52(5):1219–1231.

Heckman, J. J. (1974). Shadow prices, market wages and labor supply. *Econometrica*, 42(4):679–694.

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

Hoffmann, R. and Kassouf, A. L. (2005). Deriving conditional and unconditional marginal effects in log earnings equations estimated by Heckman's procedure. *Applied Economics*, 37:1303–1311.

Jones, S. R. G. and Riddell, W. C. (1999). The measurement of unemployment: An empirical approach. *Econometrica*, 67(1):147–161.

Jones, S. R. G. and Riddell, W. C. (2006). Unemployment and nonemployment: heterogeneties in labor market states. *The Review of Economics and Statistics*, 88(2):314–323.

Kalb, G. R. (2002). Estimation of labour supply models for four separate groups in the Australian population. Melbourne Institute Working Paper No. 24/02. Available at `http://melbourneinstitute.com/wp/wp2002n24.pdf`.

Kalb, G. R. and Lee, W.-S. (2008). Childcare use and parents labour supply in Australia. *Australian Economic Papers*, 47(3):272–295.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica*, 51(2):507–512.

Little, R. J. A. and Su, H. L. (1989). Item non-response in panel surveys. In Kasprzyk, D., Duncan, G. J., Kalton, G., and Singh, M. P., editors, *Panel Surveys*. New York: Wiley.

McFadden, D. L. (1973). Conditional logit analysis of of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*. New York: Academic Press.

Puhani, P. A. (2000). The Heckman correction for sample selection and its critique. *Journal of Economic Surveys*, 14(1):53–68.

Rammohan, A. and Whelan, S. (2005). Child care and female employment decisions. *Australian Journal of Labour Economics*, 9(2):203–225.

Ross, R. T. (1986). Analysis of the 1980 Sydney survey of work patterns of married women: Further results. *Economic Record*, 62(178):325–337.

Schaffner, J. A. (1998). Generating conditional expectations from models with selectivity bias: comment. *Economics Letters*, 58:255–261.

Schmertmann, C. P. (1994). Selectivity bias correction methods in polychotomous sample selection models. *Journal of Econometrics*, 60:101–132.

Starick, R. and Watson, N. (2007). Evaluation of alternative income imputation methods for the HILDA survey. Technical report. HILDA Project Discussion Paper Series, No. 1/07.

Stolzenberg, R. M. and Relles, D. A. (1997). Tools for intution about sample selection bias and its corrections. *American Sociological Review*, 62(3):494–507.

Tano, D. K. (1991). Are unemployment and out of the labour force behaviorally distinct labor force states? new evidence from the gross change data. *Economics Letters*, 36.

Vella, F. (1988). Generating conditional expectations from models with selectivity bias. *Economics Letters*, 28:97–103.

Watson, N. (2004). Income and wealth imputation for waves 1 and 2. Technical report. HILDA Project Technical Paper Series, No. 3/04.

Watson, N. and Wooden, M. (2002). The Household, Income and Labour Dynamics in Australia (HILDA) survey: wave 1 survey methodology. Technical report. HILDA Project Technical Paper Series, No. 1/02.

Appendix A

Table A1: Sample sizes by wave, employment status, gender, and marital/parental status

		Wave				
		1	2	3	4	5
Married males	*E*	1802	1519	1484	1385	1418
	U	77	47	43	22	30
	M	24	31	24	24	20
	N	35	48	23	24	19
Married females	*E*	1365	1140	1114	1075	1131
	U	44	40	40	31	39
	M	205	148	133	92	86
	N	324	317	287	257	231
Single males	*E*	350	327	331	350	344
	U	40	32	24	25	17
	M	15	20	19	17	12
	N	10	19	5	8	6
Single females	*E*	288	284	289	275	284
	U	17	16	13	10	10
	M	9	6	9	10	10
	N	11	23	7	8	11
Lone parents	*E*	305	300	315	303	322
	U	37	26	24	27	33
	M	99	80	67	67	63
	N	70	76	74	75	65

E: Employed; M: Marginally attached; N: Not in the labour force; U: Unemployed

Table A2: Number of individuals entering employment by year and employment state in previous month. Average of monthly transitions over calendar year, individuals ages 25-59.

Year	**From Unemployed**	**From NILF**	**Total**
2001	59.7 (19.7%)	111.1 (7.0%)	170.8 (9.0%)
2002	61.4 (20.6%)	98.1 (6.2%)	159.6 (8.4%)
2003	60.2 (21.1%)	105.5 (6.6%)	165.7 (8.8%)
2004	56.8 (22.4%)	106.5 (6.6%)	163.3 (8.8%)
2005	56.6 (24.0%)	106.7 (6.9%)	163.3 (9.2%)
2006	57 (24.3%)	112.8 (7.4%)	169.8 (9.7%)

Source: Australian Bureau of Statistics (2007) Labour Force Survey

Table A3: Number of individuals entering employment by employment state in previous month. Average of monthly transitions from 2001-2006 by sex, individuals ages 25-59.

Subgroup	From Unemployed	From NILF	Total
Male	32.2 (22.3%)	34 (8.2%)	66.2 (11.8%)
Female	26.4 (21.4%)	72.8 (6.3%)	99.1 (7.7%)
Total	58.6 (21.8%)	106.8 (6.8%)	165.4 (9.0%)

Source: Australian Bureau of Statistics (2007) Labour Force Survey

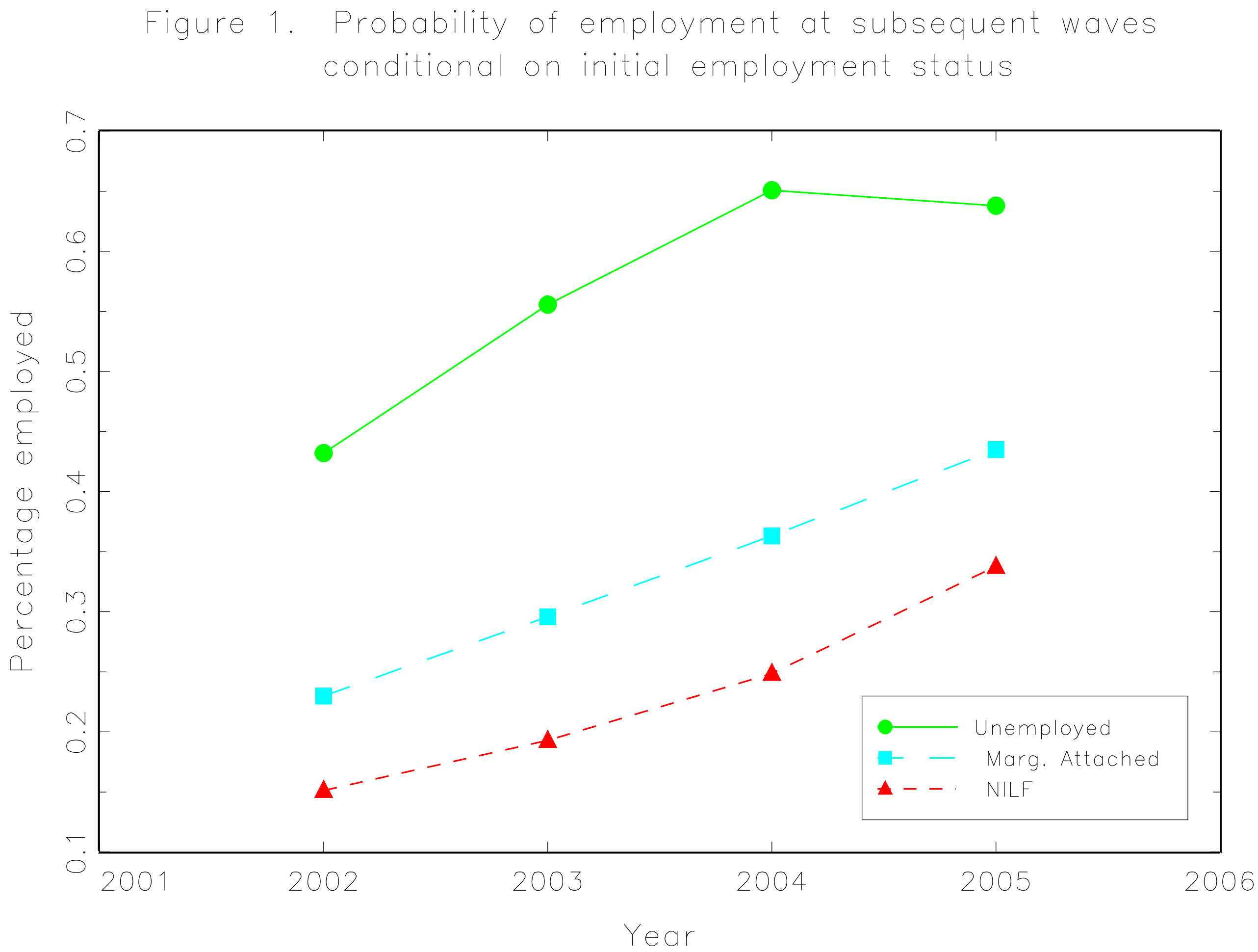
Table A4: Sign and significance of Heckman correction term in models of Tables 9 to 14

Wave	1	2	3	4	5	Pooled 1-5	Pooled 2-5
Married women							
Whole sample (E,U,M,N) Tables 9,10,11	$+^{***}$	$+^{**}$	$+$	$+$	$+$	$+^{***}$	$+^{***}$
Reduced Sample (E,U) Table 9	$-$	$-$	$+$	$-$	$-$	$-^{***}$	$-^{**}$
Reduced Sample (E,M) Table 10	$+^{***}$	$+$	$+$	$-$	$+$	$+^{***}$	$+$
Reduced Sample (E,N) Table 11	$+^{***}$	$+^{***}$	$+$	$+$	$+^{**}$	$+^{***}$	$+^{***}$
Lone Parents							
Whole sample (E,U,M,N) Tables 12,13,14	$-$	$+^{**}$	$-$	$+$	$-$	$-$	$-$
Reduced Sample (E,U) Table 12	$-^{*}$	$+$	$-$	$-$	$-$	$-$	$-$
Reduced Sample (E,M) Table 13	$-$	$+$	$+$	$+$	$-^{**}$	$-$	$-$
Reduced Sample (E,N) Table 14	$-$	$+^{*}$	$-$	$+$	$-$	$-$	$+$

*** indicates significance at the 1% level.
** and $*$ indicate significance at the 5% and 10% levels, respectively.

Table A5: Descriptive statistics by population sub-group
Wave 5 averages and standard deviations

Subgroup	**Married males**	**Married females**	**Single males**	**Single females**	**Lone parents**
Observations	1487	1487	379	315	483
Proportion working	0.954 (0.21)	0.761 (0.427)	0.908 (0.29)	0.902 (0.298)	0.667 (0.472)
$\frac{\text{age}}{100}$	0.416 (0.087)	0.395 (0.084)	0.398 (0.093)	0.43 (0.107)	0.42 (0.082)
$\left(\frac{\text{age}}{100}\right)^2$	0.181 (0.073)	0.163 (0.068)	0.167 (0.076)	0.196 (0.09)	0.183 (0.068)
poorenglish	0.008 (0.089)	0.012 (0.109)	0 (0)	0.003 (0.056)	0.017 (0.128)
nsw	0.305 (0.46)	0.305 (0.46)	0.28 (0.449)	0.26 (0.44)	0.302 (0.46)
capitalcity	0.633 (0.482)	0.633 (0.482)	0.652 (0.477)	0.692 (0.462)	0.594 (0.492)
university degree	0.298 (0.457)	0.319 (0.466)	0.23 (0.421)	0.375 (0.485)	0.203 (0.403)
trade, diploma, or certificate	0.426 (0.495)	0.247 (0.431)	0.414 (0.493)	0.308 (0.462)	0.35 (0.477)
less than year 12 schooling	0.179 (0.383)	0.28 (0.449)	0.24 (0.428)	0.203 (0.403)	0.321 (0.467)
experience	22.961 (9.581)	16.763 (8.849)	19.913 (10.44)	20.451 (10.87)	17.538 (10.388)
$\frac{\text{experience}^2}{100}$	6.189 (4.542)	3.593 (3.395)	5.052 (4.638)	5.36 (4.631)	4.153 (3.877)
$\frac{\text{partner's wage}}{100}$	5.438 (4.786)	11.842 (7.24)	n/a	n/a	n/a
$\frac{\text{unearned income}}{1000}$	4.174 (24.02)	4.174 (24.02)	2.887 (10.268)	2.391 (19.809)	4.301 (14.321)
Resident children aged					
0-4 years	0.258 (0.437)	0.26 (0.439)	n/a	n/a	0.164 (0.37)
5-14 years	0.41 (0.492)	0.455 (0.498)	n/a	n/a	0.619 (0.486)
15-24 years	0.217 (0.412)	0.241 (0.428)	n/a	n/a	0.476 (0.5)
has non-resident children	0.191 (0.393)	0.147 (0.354)	0.343 (0.475)	0.149 (0.357)	0.255 (0.436)
public tenant	0.009 (0.093)	0.009 (0.093)	0.021 (0.144)	0.044 (0.206)	0.104 (0.305)
outright home owner	0.247 (0.432)	0.247 (0.432)	0.158 (0.366)	0.235 (0.425)	0.172 (0.378)
=1 if variable is imputed					
unearned income	0.072 (0.259)	0.069 (0.254)	0.095 (0.294)	0.111 (0.315)	0.118 (0.323)
wage	0.02 (0.141)	0.017 (0.129)	0.018 (0.135)	0.041 (0.199)	0.017 (0.128)
partner's wage	0.017 (0.129)	0.02 (0.141)	n/a	n/a	n/a
male	1	0	1	0	0.124 (0.33)



Data Source: Analysis sample from HILDA (see section three)

Appendix B

Table B1: Married Men
Can we pool the unemployed, the marginally attached, and not in the labour force?
Table contains p-values for test of equality of labour force states

Test[a]	Wave 1	2	3	4	5	Pooled
Can we pool the not in the labour force and the unemployed?						
T1	.34	.00**	.86	.58	.19	.01**
T2	.34	.00**	.04**	.00**	.15	.00**
T3	.27	.00**	.92	.84	.29	.00**
T4	.43	.00**	.71	.09*	.23	.00**
T5	.26	.00**	.89	.63	.26	.02**
Can we pool the marginally attached and the unemployed?						
T1	.05**	.00**	.91	.79	.00**	.00**
T2	.20	.01**	.00**	.84	.14	.00**
T3	.25	.00**	.99	.99	.00**	.01**
T4	.53	.07*	.01**	.95	.74	.04**
T5	.36	.02**	.06*	.97	.00	.04**
Can we pool the marginally attached and the not in the labour force ?						
T1	.34	.05**	.29	.60	.00**	.00**
T2	.02**	.22	.01**	.97	.42	.00**
T3	.50	.55	.39	.89	.01**	.03**
T4	.60	.68	.06*	.94	.94	.02**
T5	.64	.57	.15	.91	.01**	.06*
Can we pool the not in the labour force (including the marginally attached) and the unemployed?						
T1	.05**	.00**	.91	.56	.02**	.00**
T2	.14	.00**	.05**	.06*	.05*	.00**
T3	.13	.00**	.95	.94	.06*	.00**
T4	.24	.00**	.75	.41	.07*	.01**
T5	.16	.00**	.91	.74	.06*	.00**

Null hypothesis is that the two labour force states can be pooled. Table contains p-values.
[a] The five hypothesis tests, T1-T5 are described in detail in the text.

Table B2: Single men and women
Can we pool the unemployed, the marginally attached, and not in the labour force?
Table contains p-values for test of equality of labour force states

	Wave					
Test[a]	1	2	3	4	5	Pooled
Can we pool the not in the labour force and the unemployed?						
T1	.07*	.00**	.99	.02**	.52	.00**
T2	.69	.01**	.40	.00**	.13	.00**
T3	.21	.01**	.99	.24	.60	.00**
T4	.72	.03**	.65	n/a	.59	.00**
T5	.48	.01**	.57	.34	.52	.00**
Can we pool the marginally attached and the unemployed?						
T1	.74	.04**	.50	.24	.67	.01**
T2	.49	.03**	.12	.40	.53	.00**
T3	.40	.05**	.56	.60	.74	.03**
T4	.51	.38	.56	.72	.84	.03**
T5	.47	.19	.55	.66	.76	.02**
Can we pool the marginally attached and the not in the labour force ?						
T1	.88	.11	.33	.00**	.07*	.00**
T2	.40	.24	.17	.00**	.32	.00**
T3	.91	.41	.59	.00*	.35	.08*
T4	.83	.44	.39	.13	.61	.10*
T5	.87	.38	.65	.10*	.40	.08*
Can we pool the not in the labour force (including the marginally attached) and the unemployed?						
T1	.21	.00**	.03**	.02**	.67	.00**
T2	.68	.00**	.48	.80	.63	.00**
T3	.99	.01**	.41	.25	.77	.00**
T4	.37	.02**	.61	.92	.81	.00**
T5	.24	.01**	.53	.86	.75	.00**

Null hypothesis is that the two labour force states can be pooled. Table contains p-values.
[a] The five hypothesis tests, T1-T5 are described in detail in the text.

Table B3: Observed and predicted wages
Married women who transition from Unemployed and Marginally Attached (U+M) to employed

	From wave:	1	2	3	4	Pooled	Pooled
	To wave:	2	3	4	5	All waves	W2-W5 only
	Observed in data	20.26	16.57	18.11	18.02	18.42	17.65
	Predictions from different models						
M1	**Linear regression**	19.08 (0.61)	18.75* (0.07)	19.14 (0.41)	21.45** (0.03)	19.82** (0.03)	20.21*** (0.00)
Heckman selection model using whole sample							
M2	**Conditional predictor**	14.11*** (0.00)	15.90 (0.14)	17.61 (0.44)	20.12 (0.21)	16.49*** (0.00)	17.84 (0.43)
M3	**Unconditional predictor**	18.27** (0.02)	18.37 (0.37)	18.93 (0.74)	21.32* (0.05)	19.35 (0.77)	19.89* (0.06)
Heckman selection model using employed, unemployed and marginally attached							
M4	**Conditional predictor**	11.33*** (0.00)	15.39* (0.10)	15.71** (0.01)	25.49*** (0.00)	15.75*** (0.00)	19.24* (0.09)
M5	**Unconditional predictor**	18.16** (0.02)	18.49 (0.19)	18.82 (0.84)	21.64** (0.01)	19.49 (0.30)	20.14*** (0.01)

Table B4: Observed and predicted wages
Married women who transition from Not in the labour force (N+M) to employed

	From wave:	1	2	3	4	Pooled	Pooled
	To wave:	2	3	4	5	All waves	W2-W5 only
	Observed in data	18.79	17.32	21.62	19.80	19.43	19.69
	Predictions from different models						
M1	**Linear regression**	18.49 (0.64)	18.77* (0.08)	19.27 (0.17)	21.70* (0.06)	19.70 (0.26)	20.26 (0.12)
Heckman selection model using whole sample							
M2	**Conditional predictor**	14.06*** (0.00)	16.05* (0.07)	17.91*** (0.00)	20.44 (0.40)	16.62*** (0.00)	18.04*** (0.01)
M3	**Unconditional predictor**	17.73*** (0.00)	18.41 (0.53)	19.06* (0.05)	21.58 (0.10)	19.25 (0.23)	19.96 (0.77)
Heckman selection model using employed, not in the labour force and marginally attached							
M4	**Conditional predictor**	14.03*** (0.00)	15.92* (0.05)	18.05*** (0.00)	20.26 (0.48)	16.45*** (0.00)	17.84*** (0.00)
M5	**Unconditional predictor**	17.77*** (0.01)	18.42 (0.50)	19.09* (0.07)	21.57 (0.11)	19.25 (0.25)	19.95 (0.77)

Table B5: Predicted wages
Lone parents who transition from Unemployed and Marginally attached (U+M) to employed

		1	2	3	4	Pooled	Pooled
	From wave:	1	2	3	4	Pooled	Pooled
	To wave:	2	3	4	5	All waves	W2-W5 only
	Observed in data	14.35	12.37	16.20	20.80	16.03	16.56
Predictions from different models							
M1	**Linear regression**	16.05 (0.28)	18.37** (0.02)	19.35* (0.07)	19.41 (0.70)	18.15** (0.02)	18.90** (0.03)
Heckman selection model using whole sample							
M2	**Conditional predictor**	18.80*** (0.00)	13.12 (0.80)	20.19** (0.02)	17.39 (0.13)	19.36*** (0.00)	19.71*** (0.01)
M3	**Unconditional predictor**	16.88*** (0.01)	16.90 (0.14)	19.52** (0.04)	18.63 (0.32)	18.46*** (0.00)	19.10*** (0.01)
Heckman selection model using employed, unemployed and marginally attached							
M4	**Conditional predictor**	20.69*** (0.00)	13.37 (0.60)	20.18** (0.02)	17.49 (0.16)	19.14*** (0.00)	21.22*** (0.00)
M5	**Unconditional predictor**	17.24*** (0.00)	17.39* (0.07)	19.45** (0.05)	18.78 (0.41)	18.68*** (0.00)	19.33*** (0.01)

Table B6: Predicted wages
Lone parents who transition from Not in the labour force and Marginally attached (N+M) to employed

	From wave:	1	2	3	4	Pooled	Pooled
	To wave:	2	3	4	5	All waves	W2-W5 only
	Observed in data	14.64	14.25	16.00	20.00	16.55	17.13
Predictions from different models							
M1	**Linear regression**	16.67 (0.40)	18.66* (0.09)	19.43* (0.09)	19.98 (0.62)	18.49* (0.09)	19.20 (0.11)
Heckman selection model using whole sample							
M2	**Conditional predictor**	19.64*** (0.00)	13.34 (0.52)	20.22** (0.04)	17.93** (0.04)	19.73*** (0.00)	20.02** (0.03)
M3	**Unconditional predictor**	17.62** (0.03)	17.23 (0.47)	19.55* (0.07)	19.35 (0.21)	18.77** (0.02)	19.38* (0.07)
Heckman selection model using employed, not in the labour force and marginally attached							
M4	**Conditional predictor**	19.06*** (0.01)	12.84 (0.40)	20.00** (0.05)	17.53** (0.03)	18.86*** (0.01)	19.66** (0.05)
M5	**Unconditional predictor**	17.36* (0.07)	17.15 (0.48)	19.50* (0.07)	19.33 (0.20)	18.67** (0.04)	19.29* (0.09)

Table B7: Sign and significance of Heckman correction term in models of Tables B3 to B6

Wave	1	2	3	4	5	Pooled 1-5	Pooled 2-5
Married women							
Reduced sample (E,M,U) Table B3	$+^{***}$	$+$	$+$	$-$	$-$	$+^{***}$	$+$
Reduced Sample (E,M,N) Table B4	$+^{***}$	$+^{***}$	$+$	$+$	$+$	$+^{***}$	$+^{***}$
Lone Parents							
Reduced sample (E,M,U) Table B5	$-$	$+$	$-$	$+$	$-^{*}$	$-$	$-$
Reduced Sample (E,M,N) Table B6	$-$	$+^{**}$	$-$	$+$	$-$	$-$	$-$

*** indicates significance at the 1% level.
** and $*$ indicate significance at the 5% and 10% levels, respectively.

Table B8: Descriptive statistics by population sub-group
Wave 4 averages and standard deviations

Subgroup	Married males	Married females	Single males	Single females	Lone parents
Observations	1487	1487	379	315	483
Proportion working	0.954 (0.21)	0.761 (0.427)	0.908 (0.29)	0.902 (0.298)	0.667 (0.472)
age / 100	0.416 (0.087)	0.395 (0.084)	0.398 (0.093)	0.43 (0.107)	0.42 (0.082)
square (age / 100)	0.181 (0.073)	0.163 (0.068)	0.167 (0.076)	0.196 (0.09)	0.183 (0.068)
poorenglish	0.008 (0.089)	0.012 (0.109)	0 (0)	0.003 (0.056)	0.017 (0.128)
nsw	0.305 (0.46)	0.305 (0.46)	0.28 (0.449)	0.26 (0.44)	0.302 (0.46)
capitalcity	0.633 (0.482)	0.633 (0.482)	0.652 (0.477)	0.692 (0.462)	0.594 (0.492)
university degree	0.298 (0.457)	0.319 (0.466)	0.23 (0.421)	0.375 (0.485)	0.203 (0.403)
trade, diploma	0.426 (0.495)	0.247 (0.431)	0.414 (0.493)	0.308 (0.462)	0.35 (0.477)
less than year 12 schooling	0.179 (0.383)	0.28 (0.449)	0.24 (0.428)	0.203 (0.403)	0.321 (0.467)
experience	22.961 (9.581)	16.763 (8.849)	19.913 (10.44)	20.451 (10.87)	17.538 (10.388)
experience squared / 100	6.189 (4.542)	3.593 (3.395)	5.052 (4.638)	5.36 (4.631)	4.153 (3.877)
partner's wage / 100	5.438 (4.786)	11.842 (7.24)	*na* (*na*)	*na* (*na*)	*na* (*na*)
unearned income / 1000	4.174 (24.02)	4.174 (24.02)	2.887 (10.268)	2.391 (19.809)	4.301 (14.321)
resident children aged 0-4 years	0.258 (0.437)	0.26 (0.439)	*na* (*na*)	*na* (*na*)	0.164 (0.37)
resident children aged 5-14 years	0.41 (0.492)	0.455 (0.498)	*na* (*na*)	*na* (*na*)	0.619 (0.486)
resident children aged 15-24 years	0.217 (0.412)	0.241 (0.428)	*na* (*na*)	*na* (*na*)	0.476 (0.5)
has non-resident children	0.191 (0.393)	0.147 (0.354)	0.343 (0.475)	0.149 (0.357)	0.255 (0.436)
public tenant	0.009 (0.093)	0.009 (0.093)	0.021 (0.144)	0.044 (0.206)	0.104 (0.305)
outright home owner	0.247 (0.432)	0.247 (0.432)	0.158 (0.366)	0.235 (0.425)	0.172 (0.378)
has imputed unearned income	0.072 (0.259)	0.069 (0.254)	0.095 (0.294)	0.111 (0.315)	0.118 (0.323)
has imputed wage	0.02 (0.141)	0.017 (0.129)	0.018 (0.135)	0.041 (0.199)	0.017 (0.128)
has partner's wage been imputed	0.017 (0.129)	0.02 (0.141)	*na* (*na*)	*na* (*na*)	*na* (*na*)
male	1 (0)	0 (0)	1 (0)	0 (0)	0.124 (0.33)
Not in the labour force	19	231	6	11	65
Unemployed	30	39	17	10	33
Working part-time	65	547	33	67	142
Working full-time	1353	584	311	217	180
Marginally attached	20	86	12	10	63

Table B9: Descriptive statistics by population sub-group
Wave 3 averages and standard deviations

Subgroup	**Married males**	**Married females**	**Single males**	**Single females**	**Lone parents**
Observations	1487	1487	379	315	483
Proportion working	0.954 (0.21)	0.761 (0.427)	0.908 (0.29)	0.902 (0.298)	0.667 (0.472)
age / 100	0.416 (0.087)	0.395 (0.084)	0.398 (0.093)	0.43 (0.107)	0.42 (0.082)
square (age / 100)	0.181 (0.073)	0.163 (0.068)	0.167 (0.076)	0.196 (0.09)	0.183 (0.068)
poorenglish	0.008 (0.089)	0.012 (0.109)	0 (0)	0.003 (0.056)	0.017 (0.128)
nsw	0.305 (0.46)	0.305 (0.46)	0.28 (0.449)	0.26 (0.44)	0.302 (0.46)
capitalcity	0.633 (0.482)	0.633 (0.482)	0.652 (0.477)	0.692 (0.462)	0.594 (0.492)
university degree	0.298 (0.457)	0.319 (0.466)	0.23 (0.421)	0.375 (0.485)	0.203 (0.403)
trade, diploma	0.426 (0.495)	0.247 (0.431)	0.414 (0.493)	0.308 (0.462)	0.35 (0.477)
less than year 12 schooling	0.179 (0.383)	0.28 (0.449)	0.24 (0.428)	0.203 (0.403)	0.321 (0.467)
experience	22.961 (9.581)	16.763 (8.849)	19.913 (10.44)	20.451 (10.87)	17.538 (10.388)
experience squared / 100	6.189 (4.542)	3.593 (3.395)	5.052 (4.638)	5.36 (4.631)	4.153 (3.877)
partner's wage / 100	5.438 (4.786)	11.842 (7.24)	*na* (*na*)	*na* (*na*)	*na* (*na*)
unearned income / 1000	4.174 (24.02)	4.174 (24.02)	2.887 (10.268)	2.391 (19.809)	4.301 (14.321)
resident children aged 0-4 years	0.258 (0.437)	0.26 (0.439)	*na* (*na*)	*na* (*na*)	0.164 (0.37)
resident children aged 5-14 years	0.41 (0.492)	0.455 (0.498)	*na* (*na*)	*na* (*na*)	0.619 (0.486)
resident children aged 15-24 years	0.217 (0.412)	0.241 (0.428)	*na* (*na*)	*na* (*na*)	0.476 (0.5)
has non-resident children	0.191 (0.393)	0.147 (0.354)	0.343 (0.475)	0.149 (0.357)	0.255 (0.436)
public tenant	0.009 (0.093)	0.009 (0.093)	0.021 (0.144)	0.044 (0.206)	0.104 (0.305)
outright home owner	0.247 (0.432)	0.247 (0.432)	0.158 (0.366)	0.235 (0.425)	0.172 (0.378)
has imputed unearned income	0.072 (0.259)	0.069 (0.254)	0.095 (0.294)	0.111 (0.315)	0.118 (0.323)
has imputed wage	0.02 (0.141)	0.017 (0.129)	0.018 (0.135)	0.041 (0.199)	0.017 (0.128)
has partner's wage been imputed	0.017 (0.129)	0.02 (0.141)	*na* (*na*)	*na* (*na*)	*na* (*na*)
male	1 (0)	0 (0)	1 (0)	0 (0)	0.124 (0.33)
Not in the labour force	19	231	6	11	65
Unemployed	30	39	17	10	33
Working part-time	65	547	33	67	142
Working full-time	1353	584	311	217	180
Marginally attached	20	86	12	10	63

Table B10: Descriptive statistics by population sub-group
Wave 2 averages and standard deviations

Subgroup	**Married males**	**Married females**	**Single males**	**Single females**	**Lone parents**
Observations	1487	1487	379	315	483
Proportion working	0.954 (0.21)	0.761 (0.427)	0.908 (0.29)	0.902 (0.298)	0.667 (0.472)
age / 100	0.416 (0.087)	0.395 (0.084)	0.398 (0.093)	0.43 (0.107)	0.42 (0.082)
square (age / 100)	0.181 (0.073)	0.163 (0.068)	0.167 (0.076)	0.196 (0.09)	0.183 (0.068)
poorenglish	0.008 (0.089)	0.012 (0.109)	0 (0)	0.003 (0.056)	0.017 (0.128)
nsw	0.305 (0.46)	0.305 (0.46)	0.28 (0.449)	0.26 (0.44)	0.302 (0.46)
capitalcity	0.633 (0.482)	0.633 (0.482)	0.652 (0.477)	0.692 (0.462)	0.594 (0.492)
university degree	0.298 (0.457)	0.319 (0.466)	0.23 (0.421)	0.375 (0.485)	0.203 (0.403)
trade, diploma	0.426 (0.495)	0.247 (0.431)	0.414 (0.493)	0.308 (0.462)	0.35 (0.477)
less than year 12 schooling	0.179 (0.383)	0.28 (0.449)	0.24 (0.428)	0.203 (0.403)	0.321 (0.467)
experience	22.961 (9.581)	16.763 (8.849)	19.913 (10.44)	20.451 (10.87)	17.538 (10.388)
experience squared / 100	6.189 (4.542)	3.593 (3.395)	5.052 (4.638)	5.36 (4.631)	4.153 (3.877)
partner's wage / 100	5.438 (4.786)	11.842 (7.24)	*na* (*na*)	*na* (*na*)	*na* (*na*)
unearned income / 1000	4.174 (24.02)	4.174 (24.02)	2.887 (10.268)	2.391 (19.809)	4.301 (14.321)
resident children aged 0-4 years	0.258 (0.437)	0.26 (0.439)	*na* (*na*)	*na* (*na*)	0.164 (0.37)
resident children aged 5-14 years	0.41 (0.492)	0.455 (0.498)	*na* (*na*)	*na* (*na*)	0.619 (0.486)
resident children aged 15-24 years	0.217 (0.412)	0.241 (0.428)	*na* (*na*)	*na* (*na*)	0.476 (0.5)
has non-resident children	0.191 (0.393)	0.147 (0.354)	0.343 (0.475)	0.149 (0.357)	0.255 (0.436)
public tenant	0.009 (0.093)	0.009 (0.093)	0.021 (0.144)	0.044 (0.206)	0.104 (0.305)
outright home owner	0.247 (0.432)	0.247 (0.432)	0.158 (0.366)	0.235 (0.425)	0.172 (0.378)
has imputed unearned income	0.072 (0.259)	0.069 (0.254)	0.095 (0.294)	0.111 (0.315)	0.118 (0.323)
has imputed wage	0.02 (0.141)	0.017 (0.129)	0.018 (0.135)	0.041 (0.199)	0.017 (0.128)
has partner's wage been imputed	0.017 (0.129)	0.02 (0.141)	*na* (*na*)	*na* (*na*)	*na* (*na*)
male	1 (0)	0 (0)	1 (0)	0 (0)	0.124 (0.33)
Not in the labour force	19	231	6	11	65
Unemployed	30	39	17	10	33
Working part-time	65	547	33	67	142
Working full-time	1353	584	311	217	180
Marginally attached	20	86	12	10	63

Table B11: Descriptive statistics by population sub-group
Wave 1 averages and standard deviations

Subgroup	**Married males**	**Married females**	**Single males**	**Single females**	**Lone parents**
Observations	1487	1487	379	315	483
Proportion working	0.954 (0.21)	0.761 (0.427)	0.908 (0.29)	0.902 (0.298)	0.667 (0.472)
age / 100	0.416 (0.087)	0.395 (0.084)	0.398 (0.093)	0.43 (0.107)	0.42 (0.082)
square (age / 100)	0.181 (0.073)	0.163 (0.068)	0.167 (0.076)	0.196 (0.09)	0.183 (0.068)
poorenglish	0.008 (0.089)	0.012 (0.109)	0 (0)	0.003 (0.056)	0.017 (0.128)
nsw	0.305 (0.46)	0.305 (0.46)	0.28 (0.449)	0.26 (0.44)	0.302 (0.46)
capitalcity	0.633 (0.482)	0.633 (0.482)	0.652 (0.477)	0.692 (0.462)	0.594 (0.492)
university degree	0.298 (0.457)	0.319 (0.466)	0.23 (0.421)	0.375 (0.485)	0.203 (0.403)
trade, diploma	0.426 (0.495)	0.247 (0.431)	0.414 (0.493)	0.308 (0.462)	0.35 (0.477)
less than year 12 schooling	0.179 (0.383)	0.28 (0.449)	0.24 (0.428)	0.203 (0.403)	0.321 (0.467)
experience	22.961 (9.581)	16.763 (8.849)	19.913 (10.44)	20.451 (10.87)	17.538 (10.388)
experience squared / 100	6.189 (4.542)	3.593 (3.395)	5.052 (4.638)	5.36 (4.631)	4.153 (3.877)
partner's wage / 100	5.438 (4.786)	11.842 (7.24)	*na* (*na*)	*na* (*na*)	*na* (*na*)
unearned income / 1000	4.174 (24.02)	4.174 (24.02)	2.887 (10.268)	2.391 (19.809)	4.301 (14.321)
resident children aged 0-4 years	0.258 (0.437)	0.26 (0.439)	*na* (*na*)	*na* (*na*)	0.164 (0.37)
resident children aged 5-14 years	0.41 (0.492)	0.455 (0.498)	*na* (*na*)	*na* (*na*)	0.619 (0.486)
resident children aged 15-24 years	0.217 (0.412)	0.241 (0.428)	*na* (*na*)	*na* (*na*)	0.476 (0.5)
has non-resident children	0.191 (0.393)	0.147 (0.354)	0.343 (0.475)	0.149 (0.357)	0.255 (0.436)
public tenant	0.009 (0.093)	0.009 (0.093)	0.021 (0.144)	0.044 (0.206)	0.104 (0.305)
outright home owner	0.247 (0.432)	0.247 (0.432)	0.158 (0.366)	0.235 (0.425)	0.172 (0.378)
has imputed unearned income	0.072 (0.259)	0.069 (0.254)	0.095 (0.294)	0.111 (0.315)	0.118 (0.323)
has imputed wage	0.02 (0.141)	0.017 (0.129)	0.018 (0.135)	0.041 (0.199)	0.017 (0.128)
has partner's wage been imputed	0.017 (0.129)	0.02 (0.141)	*na* (*na*)	*na* (*na*)	*na* (*na*)
male	1 (0)	0 (0)	1 (0)	0 (0)	0.124 (0.33)
Not in the labour force	19	231	6	11	65
Unemployed	30	39	17	10	33
Working part-time	65	547	33	67	142
Working full-time	1353	584	311	217	180
Marginally attached	20	86	12	10	63