THE PERSISTENCE OF UNEMPLOYMENT AMONG AUSTRALIAN MALES

Abstract

The persistence of unemployment for Australian men is investigated using the Household Income and Labour Dynamics Australia panel data for the period 2001-09. The paper examines whether true persistence or state dependence exists among Australian men and separates the effects of individual observed and unobserved characteristics on unemployment. A random effects dynamic probit model, which takes into account the endogeneity of initial conditions, is used. Results show that true persistence of unemployment exists for Australian men. Initial conditions are also found to be significant in estimating true persistence. The paper also finds that observed individual characteristics have a larger effect of about 77 per cent on the probability of unemployment than unobserved characteristics and true persistence. Overall the paper finds that true persistence exists and initial conditions, observed and unobserved characteristics need to be considered when estimating true persistence.

1.0 Introduction

Analysis of the dynamics of unemployment usually takes two routes. The first band of literature focuses on the persistence of unemployment at the macro level while the second band looks at the dynamics of the labour market at the individual level. The macro studies examine the dynamics of the aggregate unemployment rate and one commonly studied question is why the unemployment rate is persistent around a particular level (see Fahrer & heath 1992; Debelle & Swann 1998). The micro level literature analyses persistence at the individual level. In these studies persistence is defined as the probability of an individual being unemployed at a particular time conditional, on the individual's unemployment status in the previous period. A combination of three factors can explain why an individual is unemployed at a particular time. The first two explanations relate to an individual's observed and unobserved characteristics such as education level and motivation. In the third explanation the individual might be unemployed because the individual was unemployed in the previous period. This is referred to as true state dependence or true persistence (Arulampalam 1999) Thus observed persistence can be attributed to observed characteristics, unobserved heterogeneity and true persistence.

Few papers have studied the dynamics of unemployment in the Australian labour market. Knight et al. (2000) examine the dynamics of the Australian labour market using 1985 to 1988 panel data from the Australian Longitudinal Survey (ALS). They estimate an economic model that separates the effect of state dependence and unobserved heterogeneity on long-term labour market outcomes. Their study however concentrates on Australian youth. After the Knight study Carroll (2006) explores the determinants of the duration of unemployment in Australia, using the Household Income and Labour Dynamics panel data. He finds that factors that increase wage offers such as university education and work experience decrease the duration of unemployment. He also finds that factors that determine non-wage income such as disability increase the duration of unemployment.

In the international literature a notable paper by Arurampalam (2002) examines unemployment for British men using the British Household Panel Survey (BHPS) for the period 1991 to 1997 using a random effects dynamic probit model. Arulampalam finds that true persistence of unemployment exists for British men, implying that an individual's previous unemployment status has effects on future labour market behaviour.

This paper follows Arulampalam's study and estimates the persistence of unemployment among Australian Males using nine waves of the Household Income and Labour Dynamics Australia (HILDA) panel data set from 2001 to 2009 using a random effects dynamic probit model. Specifically the paper examines whether true persistence exist among Australian men and separates the effects of individual characteristics, both observable and unobservable, on unemployment. The results show that true persistence of unemployment exists for Australian men. Observed characteristics such as education, ancestry and the health condition of an individual are found to have a larger effect on unemployment than unobserved characteristics and true persistence.

The paper proceeds with an outline of the methodology in section 2. Section 3 provides a description of the data and the variables. Section 4 reports the results from the model estimation and descriptive evidence of the existence of persistence. Section 5 concludes and offers recommendations for further study.

2.0 Methodology

This paper uses a dynamic random effects probit model to estimate the persistence of unemployment among Australian males. Specifically the paper disentangles the effects of true persistence, observed characteristics and unobserved heterogeneity on observed persistence among Australian males.

To estimate the persistence of unemployment consider the model:

$$UP_{it}^{*} = \beta X_{it}' + \gamma UP_{i,t-1} + \alpha_i + \varepsilon_{it} \qquad (1)$$

$i = 1, \ldots \ldots \ldots \ldots \ldots, N$

$t = 2, \ldots \ldots \ldots \ldots., T$

The binary observed dependent variable UP_{it} has a value of one if the Australian male is unemployed at time t and zero otherwise. UP_{it}^{*} is a latent variable indicating Australian male's tendency to be unemployed. An Australian male is observed to be unemployed when his propensity to be unemployed crosses a threshold UP_{it}^{*}>0. UP_{it-1} is the previous period's unemployment status. α_i is an individual specific and time invariant unobserved effect. The parameter α_i is assumed to be normally distributed with a zero mean and variance σ_{α}^{2} and ε_{it} is a time and individual specific disturbance term, which is assumed to be uncorrelated with X_{it} and α_i and is also serially independently distributed standard normal. Since UP_{it}^{*} is unobserved the probit model is estimated with the observed unemployment variable UP_{it} , which takes the value 1when $UP_{it}^{*} > 0$ and zero otherwise.

However two issues would emerge if equation (1) is estimated using a pooled probit model. The first problem from estimating equation (1) is that the lagged variable UP_{it-1} is correlated with the total error term $\mu_{it} = \alpha_i + \varepsilon_{it}$, which creates an issue of initial conditions. Initial conditions refer to the pre-sample information such as the employment status of the individual before the sample, UP_{i0}, and the factors that affect $UP_{i0.}$ Anderson and Hsiao (1981) show that assumptions that are placed on the initial conditions affect the consistency of panel the estimates because the time period is short in most panel data but estimates are consistent when $T \rightarrow \infty$ as is the case in most time series. They also analyse four assumptions through which the initial conditions might affect the consistency of estimates. First the initial conditions might be fixed and with a common mean and no individual effects on the starting value in which case the MLE estimator will be consistent. Second the initial conditions might depend on random individual endowments α_i. In this case the initial conditions affect the starting value but the effect of the initial conditions diminishes as $T \rightarrow \infty$. Third the initial condition might be fixed with a common mean but the conditions affect the starting value unlike the first assumption. The final assumption is that the initial conditions might not be different from the other observations. However Heckman (1981) argues that the first three assumptions are unrealistic and need to be accounted for because the initial conditions are endogenous and are also a function of past unemployment and unobserved heterogeneity. Ignoring the endogeneity of initial conditions will lead to biased estimates of γ when T is small. Since in this study T= 9 the MLE for equation (1) would be inconsistent if the initial conditions are not considered.

To address the initial conditions problem a Heckman (1981) estimator is used. The observed unemployment variable UP_{i1} is approximated by a reduced form equation given by

$$UP_{i1} = \pi' Z_{i1} + \eta_i \quad (2)$$

where Z_{i1} contains information from the first wave and η_i is correlated with α_i but uncorrelated with ε_{it} fot t > 2. The orthogonal error component η_i is linearly specified as:

$$\eta_i = \theta\alpha_i + \varepsilon_{i1} \quad (3)$$

α_i and ε_{i1} are othorgonal to one another and it is assumed that ε_{i1} has the same distribution as ε_{it} for all i and t = 2, …., T.

Combining equations (2) and (3)

$$UP_{i1} = \pi' Z_{i1} + \theta\alpha_i + \varepsilon_{i1} \quad (4)$$

The exogeneity of the initial conditions is tested with the hypothesis H_0 $\theta = 0$. If the initial conditions are exogenous the model can be treated as a simple random effects probit model.

The second issue with estimating equation (1) is that the assumption of no correlation between the observed characteristics X_{it} and the individual unobserved heterogeneity α_i might not hold. This problem emerges when some important variables such as time varying individual unobserved effects are omitted from the model leading to bias in the estimate of γ. Cameron and Triveldi (2005) show that when the important unobserved variables are omitted from the model σ_α^2 is very large relative to σ_ε^2, which leads to inconsistent estimates.

To test for the significance of unobserved heterogeneity note that the correlation between the total error terms are given by:

$$\rho = Corr(\mu_{it}, \mu_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2+\sigma_\varepsilon^2}\ , \quad t, s = 2, \dots \dots \dots, T \ and \ t \neq s \qquad (5)$$

A likelihood ratio test is used to test for the presence of unobserved heterogeneity with $H_0 \ \ \rho = 0$. If $\rho = 0$ unobserved heterogeneity is not significant in the model, hence the panel estimator and the pooled estimator will not be different.

In this paper a model that assumes no correlation between individual unobserved characteristics α_i and observed characteristics and one that assumes correlations between individual unobserved characteristics α_i and observed characteristics is estimated. The model that assumes correlation between individual unobserved characteristics α_i and observed characteristics is estimated by following Mundlak (1978) approach. Individual time varying averages of time varying variables are included in the explanatory variables. Thus the individual unobserved characteristics in equation 1 are rewritten as $\alpha_i \ = \ \bar{X}_i'\lambda + \nu_i$ where $\nu_i \sim iid.$ ν_i is also assumed to be independent of X_{it} and ε_{it} .

Maximum likelihood estimation is used to obtain consistent estimates of the persistence of unemployment, γ, using equation (1) and (4) using maximum likelihood estimation. According to Greene (2012) the Log-likelihood for the Heckman approach to estimating equations (1) and (4) is given by:

$$\ln L|\alpha_i = \sum_{i=1}^{n} ln \prod_{t=1} \{\Phi\left[(2y_{it} - 1)\langle C_{it}\left(X_{i1}'\beta + \gamma y_{i,t-1}\right) + D_{it}(X_{it}'\delta + \pi Z_i) + (1 + \lambda D_{it})\alpha_i\rangle\right]\} \qquad (6)$$

where $D_{it} = 1$ in period 1 and 0 in every other period and $C_{it} = 1 - D_{it}$ The log-likelihood is maximised with respect to $(\delta, \pi, \theta, \beta, \gamma, \sigma_\alpha)$.

To interpret the estimates of the model average partial effects are used as opposed to marginal effects calculated at the means. Under the normalisation $\sigma_\mu = 1$ (Arulampalam, 1999) the marginal effects of the explanatory variables are given by:

$$\frac{\partial[prob(y_{it}=1|x_{it})]}{\partial x_{jit}} = \frac{\partial[E(y_{it})]}{\partial x_{jit}} = \frac{\partial[\Phi(x'_{it}\beta_j)]}{\partial x_{jit}} = \phi(x'_{it}\beta)\beta_j \tag{7}$$

The average partial effects are calculated by evaluating the means of each individual separately and then averaging the means over the sample (Arulampalam, 1999). However, because of the normalization $\sigma_\mu = 1$ the random effects estimates are multiplied by $\sqrt{1-\hat{\rho}}$ before using equation (7).

To interpret the persistence of unemployment, the random effects estimates are compared with the pooled probit estimates. Observed persistence from the sample is disentangled into persistence due to, observed characteristics, unobserved characteristics and true persistence. Dividing the average partial effect value for γ by the observed persistence from the sample gives the proportion of observed persistence due to true persistence. The pooled average partial effect is divided by the observed persistence to give the proportion of observed persistence due to the combined effect of unobserved heterogeneity and true persistence. The difference between the random effects estimate and the pooled estimate is the proportion of observed persistence due to unobserved heterogeneity. Subtracting the proportion due unobserved heterogeneity and true persistence from 100 gives the proportion due to observed characteristics. However these values should be treated as approximate values because they are derived from an implicit formula. The described relationships are presented in the formula:

$$\underbrace{Observed\ persistence}_{from\ sample} = observed\ characteristics$$
$$+$$
$$\underbrace{Unoberved\ characteristics + \underbrace{True\ persistence}_{Random\ effects\ estimate}}_{Pooled\ estimate} \tag{8}$$

2.1 Variable Specification

Unemployment is specified as a function of lagged unemployment status, observed factors and unobserved heterogeneity. Lagged unemployment status is represented by an indicator of unemployment status in the previous period. The key observable variables in the model are education level, age, marital status, children, local unemployment rate, long term health condition and housing. These variables reflect individual job search intensity and retention (Arulampalam 2002).

Two aspects of human capital, education and health, are included because they are more likely to affect unemployment status. Education is modelled as a set of dummy variables: 'degree' includes males with an advanced diploma or higher qualification, 'certificate' includes males who have either certificate one, two or three, and 'year 12' covers individuals who finished year 12. Those who did not finish year 12 constitute the omitted category.

Long-term health condition includes individuals who have several forms of health conditions including disabilities. However this is limited to health conditions that affect the individual's work performance. The effect of the condition on work is self-reported and can be a source of self-justification bias. However a study by Oguzoglu (2007) using HILDA data shows that self-justification bias is not significant. Long term health condition is also entered as a dummy variable where one shows that the individual has a long term health condition that limits work. The degree of severity of the condition and the extent to which the condition affects work is not considered.

Other variables in the model include a set of age dummies that are used to control for age effects. Variables for marital status, and number of children under five years are included to control for family effects. A dummy for indigenous persons is included to estimate the effect of ancestry.

Another set of variables includes demand side factors that might affect unemployment status. This includes a measure of the local unemployment rate and whether the person lives in a major city or metropolitan area. The local unemployment rate reflects labour market conditions. Labour market conditions have an impact on opportunities available to an individual. The dummy for living in a metropolitan area accounts for differences in employment opportunities.

2.2 Proxy Variables for Correlated Unobserved Heterogeneity

To account for potential correlation between the explanatory variables and unobserved heterogeneity sample averages for time varying variables are included following the Mundlak (1978) approach. In order for the average of a variable to be included, the variable must exhibit a certain level of time variation. Individual time averages are included for marital status, English ability for individuals whose first language is not English and local unemployment rate. These variables are assumed to have a certain level of time variability. The time average of education attainment is not included because it shows little time variability. However since education attainment might exhibit some form of unobserved variability a variable to reflect the age an individual left school for the first time is included because it contains information about an individual's ability and motivation in early life (Vu 2006, p. 93). The variable is entered as a dummy variable with one indicating that the individual left school at age less than 17 years.

2.3 Proxy Variables for the Initial Conditions

Background variables are used as exclusion variables in the reduced form equation. The background variables include the father's and mother's employment status when the individual was 14 years old. A set of exclusion variables that shows the various categories of work for the father and mother are also used. The broad categories used are based on the broad work type categories used by the ABS. The categories used include management and professional, trade services, intermediate trade services and physical labour. Another exclusion variable used indicates whether the individual was born in Australia or not. The choice of exclusion variables follows those used by Arurampalam (2002). However the choice was also limited by the information available in the HILDA data.

A baseline model similar to Arurampalam's model is estimated first. This model does not include the variables to approximate the correlation between unobserved heterogeneity and the explanatory variables. Two more models are estimated; an extended model that has more variables and allows correlation between unobserved heterogeneity and the explanatory variables; and a compact model that has fewer initial conditions and only one age variable, which indicates whether the individual is less than 25 or not. These two models are used to determine the robustness of the baseline model.

This approach is different from Arurampalam's, which used different employment search criteria to test the stability of the parameters.

The next section outlines the sample selection process and some of the descriptive measures of persistence. The tables used in the next section and results of the estimation are reported in the Appendix.

3.0 Data and Sample Selection

The data are from the first nine waves of the Household Income and Labour Dynamics in Australia (HILDA) survey. The HILDA survey, which began in 2001, follows a sample of Australians every year. The survey focuses on work, income, and welfare and family issues. Members of households that participated in wave one are interviewed in subsequent waves to form the basis of the panel. New households are also allowed to enter the panel in each wave. Adults aged 15 years or older are interviewed with the interviews being one year apart. The survey interviews more than 7,000 and 12,000 randomly selected households and individuals respectively across Australia as shown in Table 1. Out of the 13,969 that participated in the first wave 9,245 were still participating in wave 9 of the survey with, a wave-to-wave retention rate of above 80 per cent.

The sample is limited to individuals aged 15 to 55 years, which represents the period when individuals are most active in the labour force. Since the main focus of the study is unemployment individuals who are not in the labour force are excluded from the sample. Successive information on an individual is required in order to examine the persistence of unemployment due to the presence of the lagged variable in the model. Thus to be in the sample individuals are required to have at least two consecutive observations. Another requirement for an individual to be included in the sample is that individuals are supposed to have only one spell of continuous observations. Therefore if an individual has multiple spells of continuous observations, the longest spell is chosen and if the spells are of equal length the first spell is chosen. Unemployment is defined using the Australian Bureau of Statistics (ABS 2007) definition. The ABS definition of unemployment uses a search criteria based on the International Labour Organisation (ILO). An individual is categorised as unemployed if the person is without work (worked less than one hour in the reference week), has been looking for work in the previous four weeks and is available to start work in the reference week. This definition of unemployment does not include measures of underemployment (Campbell 2008).

The final sample for the estimation consists of 5,385 Australian males who produce 30,741 observations. Table 2 provides details on the composition of the sample. The sample starts with 3,143 males in wave 1, of which 2,783 males remain as of wave 3 and 1642 males remain in the sample by wave 9. A couple of hundred males enter the sample in each wave from wave 1 to 9. Individuals from the first wave create 21,338 observations accounting for 69 per cent of the total observations in the sample.

The sample used in the estimation is an unbalanced panel. The unbalanced panel is preferred over a balanced panel for two main reasons; to maximise sample size and to minimise potential bias from survey attrition (Vu 2006, p. 81).

4.0 Results and Discussion

4.1 Descriptive Evidence on Persistence of Unemployment

The unemployment rates observed in the HILDA data in each wave are reported in Table 3. Over the sample period the average unemployment rate is 3.57 per cent. The unemployment rate started at almost 5 per cent in 2001 and declined to below 3 per cent in 2007 and 2008 but rose to above 3 per cent in 2009. This shows that the labour market conditions improved around 2007 and the effect of the improved conditions is also seen in the observed persistence trend.

Table 3 also reports the conditional rates of unemployment in the current year given unemployment status in the previous year. These represent the continuation rates of unemployment. Table 3 also shows that the continuation rates decline in 2007 and 2008. This is probably due to low unemployment rates in those two years. The table also shows the entry rates into unemployment measured by unemployment status in the current year conditional on the individual being employed in the previous year. The entry rates also decline from 2006 to 2008 but rise again in 2009. The difference between the continuation rate and the entry rate is a measure of observed persistence in unemployment. If there was no persistence the probability of unemployment in the current period would not depend on the previous period's unemployment status and the continuation and entry rates would be identical. Thus the difference between the two rates shows the effects of observed and unobserved heterogeneity on persistence. Persistence is observed to be on average 31 percentage points.

Table 4 reports the continuation rate, entry rate and observed persistence on each of the several observable characteristics. These characteristics include education achievement, age categories, children, and health and family characteristics. The continuation and entry rates differ for the three levels of education. The rates are lowest for individuals with a degree or above and highest for individuals that did not finish year 12. Table 4 also shows that the persistence of unemployment increases with age. The continuation and entry rates increase as the age categories increase with the highest rates for Australian males who are more than 45 years old. However males who are less than 25 years have persistence rates as high as for males over 45 years. In terms of long-term health conditions, males who have long-term health conditions that limit work have markedly higher persistence rates than males who do not have any long-term health conditions.

There are also notable differences between living in a city and not living in a city. Males that live in a city have lower persistence rates than those not living in a city. Males that do not own a house or have no mortgage have higher persistence rates than males that own a house or have a mortgage. However the relationship between ownership of a house and persistence might be endogenous because individuals who have longer spells of unemployment might not be likely to own a house. This is addressed in the model estimation through the initial conditions.

The persistence rates are also different for various family characteristics. Males that are married or are in a de facto relationship have lower persistence rates than males that are not married or in a de facto relationship. Individuals that do not have any children below 5 years old have higher persistence rates than individuals that have children below 5 years old. Of all the ancestry variables the only variable that shows differences in persistence rates is whether the individual is of indigenous or Torres Strait Islander origin or not. Indigenous people have evidently higher persistence rates than non-indigenous people. Table 4 shows that it the persistence rates are similar for males who were born in Australia and those that were not born in Australia. The rates are also similar for males whose first language is English and for those whose first language is not English. This suggests that although English ability matters in the Australian market it does not matter whether an individual's first language is English.

4.2 Results from the Baseline Model

A baseline model, an extended model and a compact model are estimated. The baseline model assumes that there is no correlation between explanatory variables and unobserved heterogeneity and closely follows Arulampalam's model. The extended model assumes that there is correlation between some explanatory variables and unobserved heterogeneity and includes individual averages of marital status, local unemployment rate and English ability. The extended model also includes other variables such as dummies for living in a city and indigenous people. The compact model has two variables for initial conditions only. These variables are the mother and father's employment status when the individual was 14. The compact model also has only one age variable that distinguishes between individuals who are less than 25 years old and those that are older. A simple pooled probit is also estimated in each case but only results for the pooled probit equivalent of the baseline model are reported.

The parameter estimates for the baseline model are reported in Table 5 and for the pooled probit in Table 6. Average Partial Effects (APEs) are also reported in Table 7. The APEs are averages of the individual marginal effects and are used to interpret the results instead of the coefficients because the actual coefficients are difficult to interpret due to their nonlinear nature (Vu 2009, p. 95).

The results from the baseline and pooled probit models show the extent of true persistence of unemployment for Australian men. Being unemployed in the previous period has a positive impact on the probability of being unemployed in the current period. Results from the pooled probit suggest that being unemployed in the previous period increases the probability of being unemployed in the current period by 7.2 percentage points. The pooled probit estimate does not take into account the endogeneity of initial conditions and unobserved heterogeneity, and hence is not a casual estimate of the true persistence. The random effects dynamic probit model shows that being unemployed in the previous period increases the probability of being unemployed by 5.8 percentage points. This estimate is lower than the one for the pooled probit by over one percentage point.

The difference between the pooled probit and the random effects model provide some useful analysis of persistence. The descriptive analysis in section 4.1 shows that average observed persistence is 31 per cent. Observed persistence has the combined effect of observed characteristics, unobserved heterogeneity and true persistence. The pooled probit estimate gives the proportion of observed characteristics in observed persistence. Using the implicit formula as presented in equation (8) the pooled estimate of 7.2 per cent implies that observed characteristics constitute 77 per cent of the observed persistence in the sample. The random effects estimate of the impact of lagged unemployment status provides the measure of true persistence since it has accounted for the impact of observed characteristics, unobserved heterogeneity and initial conditions on the observed persistence. The estimate of 5.8 percentage points implies that true persistence accounts for 19 per cent of the observed persistence and unobserved heterogeneity accounts for 4 per cent only. These results suggest that observed characteristics are more important in explaining observed persistence.

The results from the random effects model also suggest that it is important to account for unobserved heterogeneity and initial conditions. The log likelihood ratio test for the significance of unobserved heterogeneity (LR test for ρ) shows that unobserved heterogeneity is significant at 1 per cent and accounts for 34 per cent of the total unexplained variation in unemployment status. The estimate of θ is significant at 1 per cent and this implies that initial conditions are endogenous.

All the key explanatory variables show the expected effect on unemployment. Education attainment is expected to improve human capital and reduce the probability of being unemployed. Compared with males who did not finish year 12 males with a degree or higher qualification have a lower probability of being unemployed by 1.5 percentage points. The effect of having a certificate is 1.2 percentage points and finishing year 12 is 1.5 percentage points.

Health is another important aspect of human capital. Having a long-term health condition that limits work increases the probability of being unemployed by 3 percentage points. The effect of long-term health conditions is the highest among all the explanatory variables except the previous period’s unemployment status.

Males who are more than 25 years old have a lower probability of being unemployed by almost 2 percentage points. Those that own a house or have a mortgage have a lower probability of being unemployed. Males that are married or are in a de facto relationship also have a lower probability of being unemployed by 1.8 percentage points. However having children who are under 5 years old does not have a significant impact on the probability of being unemployed.

Tightness in the labour market is expected to increase the probability of being unemployed. A one-percentage point increase in the local unemployment rate is increases the probability of being unemployed for Australian males by 0.5 percentage points.

The pooled estimates and random effect estimates have similar signs and are close to each other. However the pooled estimates for the explanatory variables are slightly smaller than the random effects estimates.

The initial conditions for the random effects model are reported in Table 5. The pooled probit estimates for t=1 are equivalent to the initial conditions and are reported in table 6. Background variables related to father and mother’s employment are used to identify initial conditions. Of these exclusion variables, the mother’s occupation variables have the expected signs whereas as the father’s occupation variables do not have the expected signs. Two of the father’s occupations are significant; professional and managerial and physical labour. These occupations are positively correlated to an individual’s unemployment status. This result is in contrast to what Arurampalam (2002) findings for British males. Three mother’s occupations are significant; professional and managerial, intermediate services and physical labour. These occupations are negatively correlated with an individual’s unemployment status as expected. The Wald test rejects the null hypothesis that the coefficients are jointly significant. This means that the exclusion variables assist in identifying the initial conditions equations.

4.3 Extended and Compact Models

The extended model is estimated with extra variables over the baseline model. These variables include living in a major city, a dummy for indigenous persons, a dummy for the age that the person left school, and a dummy for whether an individual’s first language is English. The extended model also includes individual time varying averages for English ability, marital status and unemployment rate. These variables are assumed to have some reasonable time variability. The average English ability would vary for individuals whose first language is not English but would remain constant for individuals whose first language is English.

A compact model with fewer explanatory variables and initial conditions is also estimated The compact model is estimated to see the impact of using more general and fewer initial conditions on the parameter estimates. The compact model is also used to check the robustness of the parameter estimates. The compact model has got only one age variable as compared to the baseline model. The age variable in this model is a dummy for individuals who 25 years old or older. The compact model is much different from the baseline model in terms of the initial conditions. There are only two initial conditions for the compact model. These include whether the father was employed when the individual was 14 years employed and a similar condition for the mother.

The coefficient estimates for the extended and compact models are reported in Table 8, the initial conditions in Table 9 and the average partial effects in Table 10. The estimate of the effect of the previous period's unemployment status is very close to the baseline model estimate for both the extended and compact models. The extended model estimates the effect at 5.7 percentage points and the compact model estimates the effect at 5.9 percentage points. These estimates are similar to the baseline estimate of 5.8 percentage points. Notable in the extended model is the effect of being an indigenous person and the average marital status. Being an indigenous person increases the probability of being unemployed by 6 percentage points which is similar to the effect of being unemployed in the previous period. Being married or in a de facto relationship for a longer period of time reduces the probability of being unemployed by 3.4 percentage points. English ability is measured on a scale of one to four with four indicating that the person does not speak English at all. Thus the positive average partial effect of English ability implies that when an individual improves his level of English the probability of unemployment reduces by 1.9 percentage points.

Leaving school at age less than 17 years increases the probability of being unemployed by 0.8 percentage points. There is also a significant effect for English being the first language. However living in a major city is not significant. Having children who are under 5 years old is not significant just like in the baseline model. The extended model suggests that being married or in a de facto relationship does not matter but what matters is the duration of the relationship because the average marital status is significant. In the compact model individuals who are 25 years or older have lower probability of being unemployed by 2.3 percentage points. In summary extending or reducing the model seems not to affect the estimates of the coefficients.

The initial conditions for the extended and compact models look similar to the baseline model. In the extended model the same exclusion variables are significant and have the same signs as in the baseline model. In the compact model both exclusion variables are significant. But more important is that the exclusion variable for the father's employment status when the individual was 14 years old has a negative sign, which is expected. This is despite the results from the baseline model, which shows that some types of the father's employment increase the probability that an individual is unemployed. Thus in general if the father or mother was employed when the individual was 14 years old the probability of being unemployed is reduced.

There are two main limitations to the analysis in this paper. First no interaction variables are included in the analysis. Including interaction variables in the models picks up some of the unobserved heterogeneity that could have been omitted. For example the education variables could be interacted with the age to allow the possibility that the effect of education might be different across the age groups. However exclusion of the interaction variables in the model would not bias the estimates very much because unobserved heterogeneity seems to be less important than observed characteristics. Second the study did not consider the duration of the persistence of unemployment. Although this would bias the results from the model the bias would be serious if there were interaction terms in the model (Vu 2009). The problem of duration could be through the inclusion of more lagged variables of unemployment. This was not explored because the panel is short. Despite the discussed caveats the results show that true persistence of unemployment exists for Australian men. The results also suggest that observed characteristics have a greater effect on unemployment than unobserved characteristics.

5.0 Conclusions and Recommendations

This paper estimated the persistence of unemployment among Australian males using a random effects dynamic probit model. The paper specifically disentangled the effects of true persistence, observed characteristics and unobserved heterogeneity on unemployment. The results show that true persistence exists among Australian males. This finding implies that an individual's previous unemployment status affects his future employment status. The estimates of true persistence were similar for the baseline, extended and compact models. The results also indicate that observed characteristics are far more important than unobserved characteristics and true persistence in explaining unemployment status for an Australian male because 77 per cent of the observed persistence was due to observed characteristics.

All the observed characteristics showed the expected impact on unemployment status. Higher education lowers the probability of unemployment; older males have a lower probability of unemployment and being married or in a de facto relationship reduce unemployment. Individuals with long-term health conditions such as disabilities have higher probabilities of unemployment. The results also showed indigenous people have a very high probability of unemployment, which was close to the effect of true persistence.

The study has also shown that it is important to account for initial conditions when estimating the persistence of unemployment. The probability of unemployment is higher for individuals whose parents were unemployed when the individual was 14 years old. However this study also shows that although the probability of unemployment is lower for an individual whose father was unemployed when the individual was 14 years old some types of work increase the probability of unemployment. These include managerial and physical labour type of work.

This study can be extended in two ways. First the study can be extended to look at the persistence of unemployment among Australian women. Extending the study in this way would allow comparisons to be made between men and women and would inform policy makers whether different policies need to be pursued for men and women. Another way to expand the study is to analyse the persistence of underemployment or to use a different definition of unemployment other than the ABS definition used in this study. Studying underemployment would be more relevant for Australia since the levels of unemployment are considerably low compared to other countries such as the UK.

6.0 Appendix

Table 1: Hilda Survey Sample Sizes and Retention

	Sample Selection		Retention	
	Households	Persons interviewed	Previous wave retention (%)	Number of wave 1 respondents
Wave 1	7682	13969		13969
Wave 2	7245	13041	86.8	11993
Wave 3	7096	12728	90.4	11190
Wave 4	6987	12408	91.6	10565
Wave 5	7125	12759	94.4	10392
Wave 6	7139	12905	94.8	10085
Wave 7	7063	12789	94.7	9628
Wave 8	7066	12785	95.2	9354
Wave 9	7234	13301	96.3	9245

Note: Previous wave retention is the percentage of respondents in the previous wave in-scope in the current wave who were interviewed.
Source: HILDA survey report volume 7

Table 2: Individual Responses by wave

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9	Total
Wave 1	3143	3143	2783	2500	2289	2119	1938	1781	1642	21338
Wave 2		321	321	260	227	205	182	170	160	1846
Wave 3			383	383	304	266	237	225	206	2004
Wave 4				314	314	258	222	202	173	1483
Wave 5					406	406	351	326	293	1782
Wave 6						311	311	243	204	1069
Wave 7							254	254	215	723
Wave 8								248	248	496
Wave 9										
Total	3143	3464	3487	3457	3540	3565	3495	3449	3141	30741

Table 3: Incidence and Persistence in the Raw Data (2001- 2009)

	Wave 1 (2001)	Wave 2 (2002)	Wave 3 (2003)	Wave 4 (2004)	Wave 5 (2005)	Wave 6 (2006)	Wave 7 (2007)	Wave 8 (2008)	Wave 9 (2009)
Employment vs Unemployment									
Prob($unemp_t$)	4.93	4.42	3.84	3.38	3.31	3.06	2.78	2.96	3.44
Prob($unemp_t \vert unemp_{t-1}$)		40	26.6	34.5	30.9	28.4	25.3	38.1	37.8
Prob($unemp_t \vert emp_{t-1}$)		1.84	1.92	1.46	1.72	1.18	1.57	1.46	2.52
Persistence: Difference		38.2	24.7	33	29.2	27.2	23.7	36.6	35.3
Ratio		22	14	24	18	24	16	26	15

Notes: (i) Persistence: difference = Prob($unemp_t|unemp_{t-1}$) - Prob($unemp_t|emp_{t-1}$)

(ii) Persistence: ratio = Prob($unemp_t|unemp_{t-1}$)/ Prob($unemp_t|emp_{t-1}$)

Table 4: Persistence Rates Across Key Demographics

Variable	Continuation rate Prob($unemp_t$|$unemp_{t-1}$)	Entry rate Prob($unemp_t$|emp_{t-1})	Persistence
Degree	19.75	1.22	18.53
Certificate	33.33	1.4	31.93
Year 12	23.94	2.22	21.72
Below year 12	45.15	2.61	42.54
Aged 25-34	24.57	1.64	22.93
Aged 35-44	32.64	1.34	31.3
Aged 45-	40.26	1.14	39.12
Aged less than 25	33.83	3.95	29.88
Owner/mortgage	27.32	1.13	26.19
Does not own a house or a mortgage	36.73	3.04	33.69
Health limits work	46.62	3.69	42.93
Health does not limit work	30.29	1.54	28.75
Married or in a de facto relationship	24.93	1.12	23.81
Not married or in a de facto relationship	37.16	3.10	34.06
Number of children under 5	25.37	1.30	24.07
Has no children under 5	33.33	1.81	31.52
Lives in a major city	28.14	1.71	26.43
Does not live in a city	40.56	1.69	38.87
Indigenous person	52	7.35	44.65
Non indigenous	31.64	1.63	30.01
Born in Australia	32.34	1.73	30.61
Not born in Australia	34.64	1.62	33.02
English is the first language	33.12	1.7	31.42
English is not the first language	36.78	1.51	35.27

Table 5: Random Effects Dynamic Probit Model Coefficient Estimates

Variable	Unemployment 2002-2009		Initial condition 2001	
Constant	-1.988***	(0.140)	-1.469***	(0.213)
Lagged unemployment	0.757***	(0.0811)		
Demand side factor				
Local unemployment rate	0.0849***	(0.0221)	0.0516*	(0.0290)
Demographics				
Aged 25-34	-0.417***	(0.0754)	-0.483***	(0.115)
Aged 35-44	-0.444***	(0.0801)	-0.644***	(0.128)
Aged 45-	-0.382***	(0.0832)	-0.697***	(0.141)
Owner/mortgage	-0.331***	(0.0517)	-0.0734	(0.0927)
Health limits work	0.483***	(0.0709)	0.713***	(0.117)
Education				
Degree or above	-0.353***	(0.0742)	-0.605***	(0.121)
Certificate I, II, or III	-0.271***	(0.0704)	-0.355***	(0.103)
Year 12	-0.340***	(0.0799)	-0.305***	(0.103)
Family				
Married or in a de facto relationship	-0.370***	(0.0582)	-0.558***	(0.0958)
Number of children under 5	0.0265	(0.0542)	0.0673	(0.0859)
Pre-sample information				
Father's occupation at age 14				
-Professional & managerial			0.246**	(0.118)
-Trade and service			0.192	(0.132)
-Intermediate trade and service			0.198	(0.138)
-labour and physical			0.498**	(0.148)
Mother's occupation at age 14				
-Professional & managerial			-0.242**	(0.113)
-Trade and service			-0.188	(0.138)
-Intermediate trade and service			-0.376***	(0.128)
-labour and physical			-0.220*	(0.115)
Rho			0.343	(0.0408)
Theta			1.246***	(0.206)
Log likelihood			-3548.24	
Number of observations			30741	
LR test of rho = 0, χ^2 (1)			344.59***	
Wald χ^2 (11)			509.39***	

Note: ***, **, and * indicate significant at 1%, 5 % and 10% respectively.

Table 6: Pooled Probit Estimates

Unemployment	t >1		t = 1	
	Coefficients	Standard errors	Coefficients	Standard errors
Constant	-1.720***	(0.101)	-0.965***	(0.143)
Lagged unemployment	1.432***	(0.0510)		
Demand side factors				
Local unemployment rate	0.0597	(0.0174)	0.0176	(0.0215)
Demographics				
Aged 25-34	-0.272***	(0.0572)	-0.307***	(0.0838)
Aged 35-44	-0.267***	(0.0578)	-0.407***	(0.0896)
Aged 45-	-0.228***	(0.0604)	-0.453***	(0.01003)
Owner/mortgage	-0.302***	(0.0391)	-0.0238	(0.0696)
Health limits work	0.394***	(0.0541)	0.583***	(0.0854)
Education				
Degree or above	-0.238***	(0.0522)	-0.445***	(0.0858)
Certificate I, II, or III	-0.198***	(0.0501)	-0.276***	(0.0752)
Year 12	-0.263***	(0.0581)	-0.260***	(0.0760)
Family				
Married or in a de facto relationship	-0.304***	(0.0438)	-0.448***	(0.0670)
Number of children under 5	0.0424	(0.404)	0.0440	(0.0644)
Pre-sample information				
Father's occupation at age 14				
-Professional & managerial			0.157*	(0.0892)
-Trade and service			0.133	(0.0996)
-Intermediate trade and service			0.154	(0.104)
-labour and physical			0.400*	(0.111)
Mother's occupation at age 14				
-Professional & managerial			-0.221***	(0.0862)
-Trade and service			-0.208**	(0.105)
-Intermediate trade and service			-0.294***	(0.0956)
-labour and physical			-0.175**	(0.0864)

LR χ^2 (19)	1397. 29	271.75***
Pseudo R^2	0.219	0.010
Log likelihood	-2490.79	-1229.74
Number of observations	25252	5380

Note: ***, **, and * indicate significant at 1%, 5 % and 10% respectively.

Table 7: Average Partial Effects

Variable	Average partial effects (percentage points)	
	Pooled Probit	Random Effects Dynamic Probit
Lagged unemployment	7.22***	5.81***
Local unemployment rate	0.3***	0.49***
Degree	-1.2***	-1.53***
Certificate	-1.0***	-1.22***
Year 12	-1.33***	-1.45***
Aged 25-34	-1.37***	-1.75***
Aged 35-44	-1.35***	-1.85***
Aged 45-	-1.15***	-1.70***
Owner/mortgage	-1.53***	-1.67***
Health limits work	1.98***	3.04***
Married or in a de facto relationship	-1.53***	-1.85***
Number of children under 5	0.2	0.13

Note: ***, **, and * indicate significant at 1%, 5 % and 10% respectively.

Table 8: Extended and Compact Random Effects Dynamic Probit Model

Unemployment	Extended model		Compact model	
	Coefficient	Standard errors	Coefficient	Standard errors
Constant	-1.596***	(0.211)	-1.988***	(0.139)
Lagged unemployment	0.736***	(0.0824)	0.770***	(0.0809)
Demand side factors				
Local unemployment rate	0.133***	(0.0284)	0.0848	0.0221
Lives in a major city	-0.0450	(0.0586)		
Demographics				
Aged 25-34	-0.371***	(0.0759)		
Aged 35-44	-0.391***	(0.0820)		
Aged 45-	-0.322***	(0.0862)		
Aged 25 or over			-0.416***	(0.0666)
Owner/mortgage	-0.298***	(0.0519)	-0.333***	(0.0507)
Health limits work	0.487***	(0.0712)	0.486***	(0.0706)
Education				
Degree or above	-0.267***	(0.0831)	-0.347***	(0.0738)
Certificate I, II, or III	-0.232***	(0.0710)	-0.267***	(0.0701)
Year 12	-0.265***	(0.0886)	-0.334***	(0.0791)
Left school at age less than 17	0.147**	(0.0648)		
Family				
Married or in a de facto relationship	-0.119	(0.0995)	-0.364***	(0.0574)
Number of children under 5	0.0197	(0.0534)	0.0179	(0.0524)
Average marriage	-0.398***	(0.121)		
Average English language ability	0.230***	(0.0539)		
Indigenous person	0.745***	(0.136)		
Rho	0.338***	(0.0424)	0.339***	(0.0409)
Theta	1.238***	(0.213)	1.224***	(0.204)
Log likelihood	-3494.55		-3541.74	
Number of observations	30741		30741	
LR test of rho = 0, χ^2 (1)	328.93***		337.35***	
Wald χ^2 (11)	579.33***		520.02***	

Table 9: Initial Conditions for the Extended and Compact Models

Unemployment	Extended model		Compact model	
	Coefficient	Standard errors	Coefficient	Standard errors
Constant	-1.188***	(0.156)	-1.160***	(0.126)
Aged 25-34	-0.464***	(0.115)		
Aged 35-44	-0.647***	(0.127)		
Aged 45-	-0.698***	(0.141)		
Aged 25 or over			-0.441***	(0.0993)
Owner/mortgage	-0.124	(0.0885)	-0.263***	(0.0960)
Health limits work	0.697***	(0.118)	0.477***	(0.0697)
Degree or above	-0.428***	(0.132)	-0.482***	(0.127)
Certificate I, II, or III	-0.283***	(0.104)	-0.322***	(0.102)
Year 12	-0.122	(0.118)	-0.161	(0.115)
Left school at age less than 17	0.323***	(0.0979)	0.245***	(0.0916)
Married or in a de facto relationship	-0.196	(0.160)	-0.586***	(0.0934)
Number of children under 5	0.0562	(0.0868)	0.113	(0.0831)
Lives in a major city	-0.0302	(0.0817)		
Average marriage	-0.525***	(0.170)		
Average English language ability	0.194***	(0.0516)		
Indigenous	0.575***	(0.202)		
Father's occupation at age 14				
-Professional & managerial	0.281**	(0.119)		
-Trade and service	0.207	(0.133)		
-Intermediate trade and service	0.212	(0.138)		
-labour and physical	0.502***	(0.149)		
Father was employed at age 14			-0.212**	(0.0978)
Mother's occupation at age 14				
-Professional & managerial	-0.193*	(0.115)		
-Trade and service	-0.132	(0.139)		

-Intermediate trade and service	-0.335***	(0.130)		
-labour and physical	-0.156	(0.116)		
Mother was employed at age 14			-0.182*	(0.103)

Table 10: Average Partial Effects

Variable	Extended pooled probit	Extended random effects model	Compact pooled probit	Compact random effects model
Lagged unemployment	6.94	5.7	7.22	5.94
Local unemployment rate	0.53	1.12	0.3	0.47
Degree	-0.95	-1.23	-1.2	-1.5
Certificate	-0.9	-1.09	-1.0	-1.19
Year 12	-1.11	-1.20	-1.34	-1.42
Aged 25-34	-1.21	-1.64		
Aged 35-44	-1.18	-1.71		
Aged 45-	-0.97	-1.49		
Aged over 25			-1.3	-2.34
Owner/mortgage	-1.39	-1.54	-1.5	-1.67
Health limits work	2.0	3.15	2.0	3.05
Married or in a de facto relationship	-0.31	-0.59	-1.51	-1.81
Number of children under 5	0.21	0.1	0.17	0.09
Lives in a major city	-0.2	-0.23		
Indigenous person	2.52	5.95		
Left school at less than 17	0.45	0.75		
English is first language	-1.04	-1.02		
Average English ability	0.84	1.85		
Average marital status	-1.62	-3.35		
Average unemployment rate	-0.43	-0.87		

7.0 References

Australian Bureau of Statistics 2007, ‘Labour statistics: concepts, sources and methods’ Australian Bureau of Statistics, Canberra, viewed 3 August 2012, <http://www.abs.gov.au/ausstats/abs@.nsf/mf/6102.0.55.001>.

Anderson, TW & Hsiao, C 1981, ‘Estimation of dynamic models with error components’ *Journal of the American Statistical Association*, vol. 76, no. 375, pp. 598-606.

Arulampalam, W 2002, ‘State dependence in unemployment incidence: evidence for British men revisited’ IZA discussion paper no. 630, Institute for the study of labour (IZA), Bonn, viewed 31 July 2012 <http://ssrn.com/abstract=358161>.

____, Booth, AL & Taylor, MP 2000, ‘Unemployment persistence’ *Oxford Economic Papers*, no. 52, pp. 24-50

Cameron, AC & Trivedi, PK 2005, *Microeconometrics: methods and applications,* Cambridge University press, Cambridge.

Campbell, I 2008, ‘Pressing towards full employment? The persistence of underemployment in Australia’ *Journal of Australian Political Economy,* vol. 61, no. 61, pp. 156-180.

Carroll, N 2006, ‘Explaining unemployment duration in Australia’ *The Economic Record,* vol. 80, no. 258, pp. 298-314.

Debelle, G & Swann, T 1998, ‘Stylised facts of the Australian labour market’ RBA Research Discussion Paper 9804, Reserve Bank of Australia, Economic Research Department, viewed 2 August 2012, <http://www.rba.gov.au/publications/rdp/1998/9804.html>.

Fahrer, J & Heath, A 1992, ‘The evolution of employment and unemployment in Australia’ RBA Research Discussion Paper no. 9215, Reserve Bank of Australia, Economic Research Department, 2 August 2012, <http://www.rba.gov.au/publications/rdp/1992/9215.html>.

Greene, WH 2012, *Econometric analysis,* 7th edn, Pearson Education limited, Essex.

Heckman, JJ 1981, ‘The incidental parameters problem and the problem of initial condition in estimating a discrete time –discrete data stochastic process’ in CF Manski & D McFadden (eds) *Structural analysis of discrete data with econometric applications,* MIT press, Cambridge, Massachusetts, pp. 179-195.

Household, Income and Labour Dynamics in Australia survey 2012, ‘Family, incomes and jobs volume 7: a statistical report on waves 1 to 9 of the Household, Income and Labour Dynamics in Australia survey’ Melbourne Institute of Applied Economic and Social Research, University of Melbourne, viewed 20 August 2012, <http://www.melbourneinstitute.com/downloads/hilda/Stat_Report/statreport-v7-2012.pdf>.

Knights, S, Harris, M & Loundes, J 2000, ‘Dynamic relationships in the Australian labour market: heterogeneity and state dependence’ Melbourne Institute Working Paper no. 6/00, University of Melbourne, viewed 20 August 2012
<http://dtl.unimelb.edu.au/R/F1STPGDMQDCV5RD1I3K7N7I8Q5U6KFQDQIAFAPA2G3M9AYTIDJ-01210?func=dbin-jump-full&object_id=65819&pds_handle=GUEST>.

Mavromaras, K, &Mcguiness, 2012, ‘Overskilling dynamics and education pathways’ *Economics of Education Review*, vol. 31, no. 5, pp. 619-628.

Mundlak, Y 1978, ‘On the pooling of time series and cross section data’ *Econometrica*, vol. 46, no. 1, pp. 69-85.

Oguzoglu, U 2007, ‘Dynamics of work limitation and work in Australia’ IZA discussion paper no. 2867, Institute for the study of labour (IZA), Bonn, viewed 12 October 2012, <http://ftp.iza.org/dp2867.pdf>.

Vu, HTH 2009, ‘The dynamics of income support in Australia’ Melbourne Institute of Applied Economic and Social Research, PhD thesis, University of Melbourne.