Dynamics of Female Labour Supply: Evidence of Differences in Preference between Single and Couple Females

A Paper Submitted to the 2009
HILDA Survey Research Conference
16th & 17th July 2009 University of Melbourne

Darcy Fitzpatrick & Laurence Lester
National Institute of Labour Studies
26 February 2009

Contact:
Darcy Fitzpatrick
National Institute of Labour Studies (NILS)
Flinders University
Adelaide, SA
Ph: 08 8201 2396
Email: Darcy.Fitzpatrick@flinders.edu.au

Introduction

This Paper presents the results of econometric modelling of female labour force participation and the supply of hours (contingent on being employed) for females in Australia based on the first six waves of the HILDA data. (The Household, Income and Labour Dynamics in Australia (HILDA) survey (funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs) is designed and managed by the Melbourne Institute of Applied Economic and Social Research at the University of Melbourne.)

We use sophisticated panel (longitudinal) data econometric models that are based on an extensive review of the recent theoretical and applied econometric literature addressing labour supply for single and partnered individuals.[1] Applied econometric models of labour force participation and hours of labour supply in this Paper:

a) investigate the set of factors which influence women's decisions, the relative importance of explanatory factors, and implied semi-elasticities (i.e. the percent change in the dependent variable for a one unit change in an explanatory variable);

b) control for unobserved individual level attributes or characteristics (i.e. unobserved heterogeneity);

c) incorporate dynamics to control for the influence of previous period values and *state dependency*[2] on the current value of the dependent variable;

d) adopt a two-stage selection model to account for potential bias in econometric estimates due to "selection bias" in models of hours of labour supply (i.e. labour supply is contingent on a labour force participant female being employed); and

e) analyses separate models for single females and for females with male partners.

Following the results of econometric modelling, we canvas the implications of the econometric model results for influencing the labour supply of women.

[1] See Chiappori (1988); Chiappori (1992); Nijman and Verbeek (1992); Fortin and Lacroix (1997); Aronsson et al. (1999); Vella and Verbeek (1999); Ligon (2002); Donnie (2003); Bloemen (2004); Chiappori and Donni (2005); Breunig et al. (2005); Creedy and Kalb (2005); Vermeulen (2005); Vermeulen (2006); Blundell et al. (2007); Couprie (2007); and van Klaveren (2008).

[2] Specifically, when correlation between observations over time (e.g. waves of panel data) is due to a mechanism influenced by the individual's state prior to the observed data.

Background

Labour Supply Shortages

Although the Global Financial Crisis (GFC) of 2008-2009 has immediate implications for previously reported labour shortages, the impact is likely to be relatively short-lived (perhaps as short as a year to two, perhaps longer). Once the GFC has run its course, Australia will likely return to the pervious serious problem of skill shortages. Such shortages are a complex problem from the point of view of employers, and are generally a consequence of a surge in the demand for skilled workers from above average annual rate of economic growth. For example, between 1992 and 2006, real Gross Domestic Product (GDP) in Australian rose by 52 per cent (ABS (2006a)). Skill or labour shortages also reflect underlying demographic change. Thus, in more usual times than currently being experienced, the official view is a projected rapid decline in labour force growth in Australia: annual labour force growth is projected to fall from the current levels of around 1.6 per cent per annum to less than 0.6 per cent over the next 20 years (Productivity Commission (2005)). The diminution in Australian labour force growth is a result of an ageing population and the resulting fall in labour force participation: retirement rather than contraction in the number of young entrants to the labour force is the main explanation for the projected fall in Australia's labour force growth rate.

Employment Issues

The increased labour market participation of women during the last 20 years (e.g. from 61 per cent in 1988 to 65 per cent in 2007—concurrent with a fall in male participation from 78 per cent to 72 per cent), particularly those married and with children, has been one of the most significant economic and social changes of recent times. Moreover, recent growth in employment has been particularly strong in casual employment (e.g. between 1992 and 2005, nationally, there as a 19 per cent increase in casual employment for men and a 16 per cent increase for women, while part-time permanent employment for women grew by 20 per cent compared to 4 per cent for men (ABS (2006b)). A new trend has also developed—the full-time casual, but this trend has affected men more than women (e.g. between 1992 and 2005 an increase of 9 per cent for men and 5 per cent for women (ABS (2006c)).

Factors Influencing Women's Labour Supply

When considering the determinants of labour market outcomes it is possible to take advantage of the extensive, detailed, theoretical, and empirical literature and thus assemble a list of variables to be included as explanatory variables in labour market models (Winkelmann (2006)). This Paper follows the second course and uses the abundant literature to identify measures that influence the probability of being a labour market participant, of being employed, and the hours supplied (see Lester (2008) for a detailed review of the literature of factors influencing labour market outcomes[3]).

In addition to variables derived from the literature review, there remain unobserved (and generally unmeasurable) individual attributes that influence labour market outcomes, including psychological and behavioural traits, motivation, self-direction, and ambition, (Groves (2005); Isacsson (2007)). General unobserved characteristics, which bias econometric estimates based on cross-sectional (or pooled panel data[4]) data, are not usually available (and are not available in the data used in this Paper—or any other potential data set), but econometric models—discussed below—can be constructed to deal with individual unobserved heterogeneity (Lester (2007)).

Labour supply behaviour of females with male partners differs from single females. Child-rearing activity is clearly undertaken more by females than males, and this continues throughout much of the female's working age life. Moreover, beyond the age when it can be assumed children have left the family home, female labour supply remains less than males (e.g. after the first child, at age 48, hour of work per annum by women is about 35 per cent that of men's hours[5]). Female labour supply, as a household decision, favours male labour force participation due to comparative advantage (e.g. women who have exited the labour market to bear children will, on average, have less labour market experience than a similar aged male and will therefore attract a lower per hour wage rate). In addition, one view is that Australian families are "time poor" and this is particularly so for working mothers. If this is so, at some point in the wage distribution, a further increase in wage may not necessarily increase labour supply: it may allow women to reduce hours worked—i.e. a "backward bending" labour supply curve, usually associated with higher wage brackets, may operate for mothers with partners.

In addition, a greater proportion of single females work 'full-time' hours compared to partnered females, and single females work an average of 31 hours per week compared to 28 hours for couple-females. Sole parent females may also have a different pattern of hours supplied.

The empirical questions therefore are, using models of female participation and hours worked (contingent on being employed), which explanatory variables (measured attributes, characteristics, or demographic factors) are shown to have a statistically significant estimated coefficient (e.g. wage rate is expected to have a positive coefficient and therefore is associated with increased hours). In the models used in this Paper, estimated coefficients can be interpreted as semi-elasticities (i.e. if an explanatory or independent variable x increases

---

[3] Lester (2008, Ch.4) examines immigrants' labour market success, but, except for immigrant specific issues (e.g. country of education), the review is equally informative for labour market outcomes for all individuals.

[4] That is, treating waves of the panel data as if they were collected at the same time, or as a single cross-section.

[5] Apps (2007)—Original data: ABS Survey of Income and Housing 2003-04.

by 1 unit, what is the percent increase or decrease in the dependent variable such as hours supplied?).

The HILDA Data

The impetus for the HILDA survey was to trace the income, labour market, and family dynamics of the Australian population, over an extended period. The first survey was conducted in 2001, with subsequent surveys conducted annually: this Paper is based on analysis of the first six waves of data 2001 to 2006 inclusive.

The initial sample selection of the HILDA survey went to great lengths to ensure that the sample was random, that attrition of respondents from year to year was minimised, and that the survey had an indefinite life. The reference population was all Australian residents who lived in private households as their primary place of residence. The sample was selected using a stratified approach by state and by metropolitan and non-metropolitan regions. Data was, and continues to be, collected through personal interviews and through self-completion questionnaires.

In the first wave, 7683 households were selected. This resulted in a sample of 15,127 persons, age 15 years or older, eligible for interview: 13,969 individuals were successfully interviewed. Subsequent interviews for later waves were conducted one year apart.

The HILDA wave-on-wave (Australia wide) attrition rates have fallen at each wave, and falls compare well with international standards: 13.2 per cent (Wave 2), 9.6 per cent (Wave 3), 5.6 per cent 8.4 per cent (Wave 4) 5.6 per cent (Wave 5), and 5.2 per cent (Wave 6). The sample increases whenever a new household is formed when a current sample member exits a multi-person household.

For this Paper, single females are defined as females that either lived alone with or without dependent children, that lived with another family member but were not a dependent child themselves, or were unrelated to all other household members (as in a share house). Couple-females are defined as those who are married or in a *de facto* relationship to a male partner, with or without dependent children.

The specific criteria for females selected for analysis in this Paper are as follow:

- Single or partnered females with or without dependants of age 18-64 years.[6] For females in a relationship, their partner had to greater than 18 years.
- Self-employed females are excluded as the distribution of their wages differs to that for wage and salary earners (i.e. the relationship between earned income and labour supply differs). Moreover, data collected from self-employed individuals is less reliable than that from wage and salary earners—in addition to the know problems associated with self-Papered income data.
- Full-time female students under the age of 24 are excluded (classified as a dependent child).

[6] It is common in labour market studies to restrict analysis to this age group (although the sample can be restricted to those under, say, 55 if it is thought that the behaviour of those nearing retirement will differ from the younger individuals).

There are 21,688 usable observations from waves 1 to 6 of the HILDA for Australia and 15,184 usable observations for females. From this sample, 15,184 females are labour force participants, and 8,597 females supply hours of labour (see Table 1 below).

Table 1: HILDA Observations (Combined Waves 1 to 6)

		Australia
Employed		
	Single Females	5,820
	Couple Females	9,364
Hours Worked		
	Single Females	3,480
	Couple Females	5,117

Notes: (1) Sample HILDA pooled data Wave 1 to 6 (unweighted). (2) Sample is unbalanced (individuals need not be present for all waves)—there are an average of approximately 2.5 observations for each individual (with a range of 1 to 6 waves of observations).

Econometric Issues for Model Building

The panel data models in this Paper deal with two issues pertaining to females: the supply of hours (the *Hours* equation, contingent on the probability of being employed) and the decision to participate in the labour force. For convenience in discussing the rational for the econometric models, the *Hours* equation (including the influence of the probability of being *Employed* equation) is considered first. The probability of being a labour force participant, represented by the *Employed* equation, is considered second. Before outlining the specific models used in this Paper it is necessary to deal with two complexities for econometric modelling:

a) Sample selection bias—present when the sub-sample being analysed (e.g. those who supply hours of work) is a non-random selection from a larger sample (e.g. all employed female labour force participants in the specified age group).

b) Panel versus cross-sectional econometric modelling—the benefits and drawbacks of using sophisticated (and more complex) panel data models versus the more common and less sophisticated cross-sectional models.

Sample Selection Bias

Sample selection bias occurs naturally in labour supply modelling as hours worked (or wage rates) and the probability of being employed (or of being a labour force participant) are inter-related.[7] Potential bias arises from the exclusion of non-working females from the sample when estimating the hours of work equation. As the hours worked of non-working females are zero, the distribution of hours is truncated. Thus, the sample of those who do supply hours overstates the desire to supply hours of work beyond that of the population of all females of the selected age range. Econometric models that do not account for 'selection bias' may not necessarily have an error that is a mean-zero random variable in the resulting sub-sample of women who supply hours of work (it generally tends to be positive), even though it is a mean-zero random variable in the population of all females. Consequently, econometric model-based estimates of coefficients may be biased and inconsistent (i.e. amongst other things, the size and statistical significance of individual model estimates or coefficients may lead to false conclusions and poor policy prescription or advice).

Since Heckman (1978, 1979), it has been commonplace in econometric analysis to correct for sample selection bias when estimating labour supply models through a two-step procedure. In the first step, a 'reduced form', or secondary, equation is specified: for example, when modelling wage outcomes, a probability of labour force participation equation (e.g. *Employed* equation) is fitted for the complete random sample. Outcomes from the reduced form equation are then used to construct a selection bias 'correction term' that is incorporated into the second step, a 'structural', or primary, labour supply equation (i.e. hours or wage), which accounts for the non-randomness of the sub-sample and controls for selection bias.

Despite the achievements of the Heckman two-step procedure in overcoming sample selection bias, its application in empirical studies has been limited to cross-sectional data (or pooled panel data) analysis (see below). It is only recently that well developed two-step

---

[7] More specifically, the dependent variable *hours*, to be explained by regression analysis, is non-randomly selected because the probability of being employed influences the number of hours worked.

panel data procedures, similar to the Heckman two-step cross-sectional procedure, have been developed (based on innovative work by Ridder (1990); Nijman and Verbeek (1992); and Vella and Verbeek (1999)). The advanced two-step estimation procedure developed by Vella and Verbeek (1999) is adopted in this Paper to estimate labour supply models (see below).

Cross-Sectional versus Panel Data Econometric Analysis

As is well documented, the consequence of using cross-sectional (or pooled panel data) is that individuals' unobserved time-constant characteristics (or unobserved heterogeneity[8]) are not considered; unobserved heterogeneity, if present, results in inefficient econometric model estimates (with high standard errors leading to lack of statistical significance of estimated parameters). Moreover, treating panel or longitudinal data as if it were a cross-section ignores the information contained in the progress or change in measured variables, and, importantly, ignores that in panel data across-time correlations are common—autocorrelation results in inefficient parameter estimates, standard errors of the estimates are biased invalidating hypothesis tests such as t-statistics, and the R^2 (coefficient of determination[9]) is no longer reliable (Greene 2003). Moreover, the scarcity of Australian longitudinal survey data has, until recently, contributed to the restriction of cross-sectional data analysis. The HILDA survey data has provided much needed longitudinal data for Australia.

Panel data models treat the unobserved heterogeneity as a random variable: alternative assumptions are that the heterogeneity is not correlated with the (exogenous) explanatory variables (the random effects model, REM) or that there is correlation (the fixed effects model, FEM).[10] There are benefits and drawbacks of both approaches to panel data modelling—subject to much discussion in the econometric literature (see e.g. Lester, 2007 for a review). Despite the advances in panel data analysis, there are few estimators for panel models with limited dependent variables and sample selection, but the Vella and Verbeek (1999) two-step procedure deals with these matters. Moreover, their method allows the inclusion in the econometric model, specification of explanatory variables that may be correlated with the unobserved heterogeneity, and of time-invariant (or slow-changing) explanatory variables that are not usual in FEM models.[11]

[8] Note that the unobserved heterogeneity is not itself of interest in the analysis: interest is in controlling for the potential bias caused by ignoring its influence.

[9] A measure of model goodness-of-fit bounded between zero and one, where $R^2 = 1$ represents a perfect fit.

[10] Notwithstanding the confusion that may be created by the nonclamature REM and FEM, in both cases the individual unobserved heterogeneity is assumed to be a random variable. In the FEM model, heterogeneity is treated as an (estimateable) individual specific dummy variable (generally resulting in the incidental parameter problem) but in the REM, individual unobserved heterogeneity is assumed to have an empirical distribution.

[11] Since the FEM model is based on first-differences (e.g. x_{it}-x_{it-1}) time-invariant explanatory variables are "differenced" out of the models.

Unitary and Collective Models of Labour Supply

In the analysis presented in this Paper, household labour supply is estimated based on the *unitary* labour supply model. Although, as outlined below, the more recent literature suggests a *collective* model of household labour supply is more appropriate for households with two (or more) adults, currently available econometric software precludes the use of the more advanced estimators.

It is commonplace in microeconomic analysis to treat household labour supply behaviour as the utility maximisation behaviour of an individual (i.e. the household is treated as if it were an individual)—referred to in the literature as the *unitary* labour supply approach. In recent years, however, the unitary approach has been criticised at the theoretical level because it assumes that the household is characterised by a single preference or utility function. In the common unitary model, a couple in a household are treated as if they are a single unit—or, as if one individual made all the decisions concerning the labour supply provision of all household members to maximise joint utility. Hence, the unitary approach does not allow individual household members' preferences to be considered, or the intra-household distribution of welfare to be identified. In Addition, the unitary model implies that household members aggregate, or pool, their incomes so that labour supply and consumption decisions are determined only by the total exogenous income (which may include welfare payments and investment income), rather than the distribution of income across household members.

Modelling labour supply of households that includes two or more income earners (such as couple households), by application of the unitary model has come under much scrutiny both theoretically and empirically recently, and in general, the theoretical restrictions that the unitary model approach imposes are not necessarily supported by the empirical literature for households that contain more than one individual. The result of the recent evaluation of the unitary model has been the development of the *collective* approach, which considers the household members individual, but interrelated, labour supply behaviour rather than the household as if it were a single unit (Chiappori 1988; 1992).

The collective approach explicitly determines household labour supply and consumption decisions by means of the individual household members' preferences or utility functions—which allow the inclusion of the partner's welfare to be considered. In the Chiappori (1988; 1992) approach, when the preferences of one or more individuals in a couple household include concern for their own welfare and the welfare of their partner,[12] then a bargaining process dictates labour supply. Thus, in the collective model of labour supply for partner households, the interaction between household members' labour supply decisions is explicitly recognised through a sharing rule based on the division of household income between the partners. The welfare function defined in the collective model can be interpreted as a method which defines an inter-household bargaining process. Labour supply is a two-stage process, first non-labour income—a function of wage rates—is divided

[12] And, their consumption is private—e.g. individuals do not share the consumption of goods such as clothes.

between individual household members, and in the second-stage individuals decide about their labour supply conditional on their share on non-labour income.[13]

An extensive review of the econometric literature, however, indicates that the application of the two-step panel model has limitations. Thus, a comprehensive model of female labour supply (hours of work) would include both single and partnered individuals, with or without children, who do or do not participate in the labour force; which includes individual utility functions for partnered households and the rules or process for joint decisions about the supply of labour hours by each household member. At the current stage of development, however, this comprehensive, inclusive, panel data model is not available. The application of the collective approach to couple households that contain children (who, in the collective model, must be treated as a public good) are still in their infancy, as are models which include non-participants—with many models restricted to households without children where the couple are both employed (Donni (2003); Bloemen (2004); Blundell *et al.* (2007); Couprie (2007); van Klaveren (2008)).

Although variations of the collective approach can be written in equation form, no theoretical micro-econometric solution has yet been devised, and hence software to estimate the model has not been written (i.e. accessible econometrics packages do not include an appropriate econometric estimator for the collective models). Currently, restrictive estimators devised that provide accessible econometric models for simplified models only. The Vella and Verbeek (1999) model, which can be applied to single and couple females separately represents the current level of sophistication available for applied econometric panel data models of labour supply, but its theoretical base is the unitary model. Consequently, the unitary approach has been adopted in this Paper for analysis of couple households with and without children. Future research should be considered to overcome this simplification, and potential misspecification when applying the unitary model to couples.

Simultaneity and the Two-Step Selection Bias Models

The collective approach, outlined above, does not require a simultaneous equation model of female and male behaviour for couple households because the interaction is accommodated through the bargaining process—hours of labour supply are an inter-related decision of household members. In the unitary model of labour supply, one approach is to consider a simultaneous equation model for individual female and male supply in which partner's wage and hours of work (or wage per hour) appear in the hours supplied equation. The drawback of this approach (abstracting from the preference for a collective model) is that wage and hours on the right-side of equations are endogenous (they are simultaneously determined). Consequently, an instrumental variables (IV) model is required. As with all IV models, it is not clear which instruments are appropriate, and there is the persistent difficulty of finding instruments that correlate with the endogenous variable but are suitable exogenous. Further, the current specification of the two-step selection bias model does not accommodate a simultaneous equation model of female and male behaviour for couple households. The approach adopted in this Paper is to include a number of male partner

[13] A useful explanation of the assumption underpinning the bargaining process between the individuals in a household is explained by application of economic Game theory which show how the economically efficient (Nash equilibrium) can be obtained (Ligon 2002; Chiappori and Donni 2005).

characteristics which attempt to capture the 'flavour' of joint decision-making to some extent, while avoiding practical and econometric problems such as the IV approach.

The next section describes the process of applying theoretical labour supply models to observed data. Models are presented in a simplified form (for a more technically detailed expose see Vella and Verbeek 1999).

The Two-step Econometric Model of Hours of Work Supplied

Having discussed the underlying theoretical implications of estimating a labour supply model, the specification of the econometric models is now considered. In this Paper, the estimation of the labour supply model follows closely the two-step panel data procedure developed by Vella & Verbeek (1999) to overcome selection bias and endogeneity in the labour supply equation.

In the Vella and Verbeek (1999) model approach, the estimates of a structural (primary) hours worked equation [1], are obtained via a reduced form (secondary) equation [2], which determines the selection rule—the probability of being employed. Equation [3] determines when the probability of being employed is positive. Equation [4] determines (based on selection equation [3]) when labour hours supply is greater than zero—Equations [3] and [4] are referred to as the censoring and selection rules.

Two-Step Panel Data Model

Structural (primary) hours supplied equation:

$$Hours_{it}^{*} = f_1\left(X_{it}, Employed_{it}; \beta_1\right) + \mu_i + \eta_{it} \qquad [1]$$

Reduced form (secondary) employment equation:

$$Employed_{it}^{*} = f_2\left(X_{it}, Employed_{i,t-1}; \beta_2\right) + \alpha_i + \nu_{it} \qquad [2]$$

Censoring and Selection Rules

Probability of being employed equation:

$$Employed_{it} = f_3\left(Employed_{it}^{*}; \beta_3\right) \qquad [3]$$

Hours Supplied Equation:

$$\begin{aligned} Hours_{it} &= Hours_{it}^{*} && \text{if } f_4\left(Employed_{i1}, \ldots, Employed_{iT}\right) = 1 \\ Hours_{it} &= 0 && \text{if } f_4\left(Employed_{i1}, \ldots, Employed_{iT}\right) = 0 \end{aligned} \qquad [4]$$

where i are individuals (survey participants, $i = 1, \ldots, N$), t is time (or survey waves, $t = 1, \ldots, T$), and f represents functions characterised by the unknown parameters (vector) β. The x are the vector of observed individual characteristics or explanatory variables (e.g. education level, children in the household, marital status, partner's attributes, etc.), and covariates or control variables which while influential are not the subject of interest in this Paper. Random, time-invariant, individual heterogeneity are represented by μ_i and α_i and random, time-variant, individual specific, independent, effects as η_{it} and ν_{it}. Note that x need not contain identical explanatory variables across functions. Starred variables are latent (unobserved) endogenous variables (i.e. preferred hours supplied, *Hours**, and the probability of a labour force participant being employed, *Employed**)—with observed counterparts (actual hours supplied, *Hours*; and whether or not employed, *Employed*). The terms $\square_i$ and $\square_i$ represent the panel-model (random) time-invariant unobserved individual effects (heterogeneity), and η_{it} and v_{it} represent the random individual-specific time-variant effects—that are assumed independent across individuals.

To specify the *correction* terms estimated in Equation [1] to be incorporated into Equation [2], allow the error component of the secondary equations (e.g. the reduced form probability of employment) to be denoted by $u_{it} = \square_i + \square_{it}$ (i.e. a combination of individual time-invariant unobserved heterogeneity and an individual time-varying component). The time-invariant correction is approximated by the mean of the time-varying component, $\overline{u_i}$ (i.e. the average of the u_{it}[14]), and the time-varying correction is u_{it}. Thus, in the *Hours* equation, the panel-model unobserved heterogeneity ($\square_i$) and time-variant heterogeneity (η_{it}) are approximated by $\overline{u_i}$ and u_{it} respectively; as $\overline{u_i}$ and u_{it} are treated as 'data' in the *Hours* equation, their parameters (coefficients) can be estimated and if non-significant suggest no endogeneity.

The model of Equations [1] to [4] demonstrates that the determination of *Employed* (the probability of employment) is a function (f_3) of the unknown parameter vector β_3, and the function f_4 indicates that *Hours* (actual worked) is only observed for positive values of *Employed*. Thus, sample selection bias in the primary *Hours* equation is controlled for by including the selection and censoring rule from the *Employed* equation ($Employment_{it}$)—the primary *Hours* equation should not be estimated without first considering what determined its sub-sample, the reduced form *Employed* equation, or parameter estimates are potentially biased and inconsistent leading to incorrect attribution of causes of hours supplied.

Thus, the two-step model depicted above describes how to control or account for selection bias, and the inclusion of individual effects controls for heterogeneity in the panel data econometric models.

For estimation, further assumptions are made: as usual, errors are normally distributed and explanatory variables are exogenous; autocorrelation in the reduced form (secondary) *Employed* probit equation errors is inadmissible, but heteroscedasticity and/or autocorrelation in the primary *Hours* equation errors can be accommodated.

As shown in the reduced form Equation [2], the model features a potential role for dynamics (e.g. $Employment_{i,t-1}$): in addition, *state dependence* is controlled by inclusion of information on the dependent variable in the period preceding the first available data period ($Employment_{i,t=0}$).[15] The inclusion of $Employment_{t=0}$ may be endogenous due to recall problems or respondents' perceptions when concurrently Papering their current and previous behaviour at $t = 1$. The example provided by Vella and Verbeek (1999) found that while endogeneity existed, it did not affect the results in any significant way, nonetheless, they suggest potential endogeneity (due to *dynamics* and/or *state dependency*) can be controlled by including a polynomial of predicted values of the dependent variable from the *Employed* equation including in the *Hours* equation.

For the purposes of this Paper, the lagged dependent variable for the zero period was constructed using information provided by respondents in the first time period above their

[14] More generally, $\overline{u_i} = W_i^{-1}\Sigma_{t=1}^{W} u_{it}$ where W_i represents the number of waves for individual *i*.

[15] In model estimation, a single variable includes both $Employment_{i,t-1}$ and $Employment_{i,t=0}$.

experience in the previous year,[16] which controls for *state dependency* (and has the benefit of preserving observations—particularly important for the relatively small samples). Potential endogeneity was controlled by the *Employed* polynomial (in addition to the selection bias and heterogeneity correction terms in the *Hours* equation).

Empirical Equations for the Two-step Model of Hours Worked

Based on the theoretical model outlined above, the empirical equations can be specified. First, *Employed* is estimated as a limited dependent variable (probit) panel data model (see Appendix I for the probit model specification). Second, *Hours* is estimated as a panel data model of a continuous dependent variable, corrected for selection bias (i.e. only employed labour force participants that supply hours worked) by inclusion of panel data *correction terms* (the $\overline{u}_i$ and u_{it} 'data'):

$$Employed_{it}^{*} = \beta_1 x_{1,it} + \ldots + \beta_k x_{k,it} + \beta_E Employment_{i,t-1} + \alpha_i + \nu_{it} \quad [5]$$

$$Hours_{it} = \beta_1 x_{1,it} + \ldots + \beta_k x_{k,it} + f_P\left(Employment_{it}; \beta_P\right) + \overline{u}_i + u_{it} \quad [6]$$

Where the quadratic function (f_P) to control for endogeneity is defined as:

$$\begin{aligned} f_P\left(Employed_{it}; \beta_P\right) = {} & \beta_{P1} Employment_{it} + \beta_{P2} Employment_{it}^2 \\ & + \beta_{P3} Employment_{it}^3 + \beta_{P4} Employment_{it}^4 \end{aligned} \quad [7]$$

where *Hours* represents the log of hours worked in paid employment per week, x represent observed independent or explanatory variables (e.g. work experience, education, health, and marital status). f_P denotes a polynomial (of pre-specified length) with unknown coefficients (β_P) controlling for endogeneity due to dynamics, and β_E controls for *state dependence*. Note that the *Employed* equation is required in contrast to examining the strict definition of participation (i.e. including employed and unemployed), because the selection bias is due to selection into employment not selection into participation—hours supplied are not independent of selection into employment—a female who is not employed does not independently select her hours of work.

[16] When new households form, period 'zero' (the previous year) may be between waves 1 and 5.

Summary Statistics

Table 2 provides a legend of variable names and description of those included in models, and Table 3, which follows, has summary statistics for these variables.

Table 2: Explanatory Variables Used in Econometric Models

Variable Name	Description	Variables Required
lnhours	Log of weekly hours worked in paid employment	Continuous
empt & empt_lag	Employed (full-time or part-time) (and the one-period lag)	Binary state
exp	Total employment experience in years	Continuous
exp2	Total employment experience in years squared	Continuous
jbsearch	Total time out of employment in years	Continuous
jbsearch2	Total time out of employment in years squared	Continuous
age 18-24 (base)	Age of females between 18-24 years	Dummy
age 25-34	Age of females between 25-34 years	Dummy
age 35-44	Age of females between 35-44 years	Dummy
age 45-54	Age of females between 45-54 years	Dummy
age 55-64	Age of females between 55-64 years	Dummy
page 18-24 (base)	Age of male partner between 18-24 years	Dummy
page 25-34	Age of male partner between 25-34 years	Dummy
page 35-44	Age of male partner between 35-44 years	Dummy
page 45-54	Age of male partner between 45-54 years	Dummy
page 55+	Age of male partner between 55+ years	Dummy
ed 1	Highest level of education is Bachelor/Graduate Diploma/Postgraduate degree	Dummy
ed 2	Highest level of education is Advanced Diploma/Diploma	Dummy
ed 3	Highest level of education is Certificate III/IV	Dummy
ed 4	Highest level of education is Certificate I/II or Year 12	Dummy
ed 5 (base)	Highest level of education is Year 11 & below, or undetermined	Dummy
ped 1	Partner's highest level of education is Bachelor/Graduate Diploma/Postgraduate degree	Dummy
ped 2	Partner's highest level of education is Advanced Diploma/Diploma	Dummy
ped 3	Partner's highest level of education is Certificate III/IV	Dummy
ped 4	Partner's highest level of education is Certificate I/II or Year 12	Dummy
ped 5 (base)	Partner's highest level of education is Year 11 & below	Dummy
ch A	One resident dependent child between the ages 0-4 years	Dummy
ch B	One resident dependent child between the ages 5-14 years	Dummy
ch C	One resident dependent child between the ages 15-24 years	Dummy

Variable Name	Description	Variables Required
ch D	Two or more resident dependent children, where at least two are between the ages of 0-4 years and any subsequent children are greater than 0 years	Dummy
ch E	Two or more resident dependent children, where one is between the ages 0-4 years and any subsequent children are greater than 4 years	Dummy
ch F	Two or more resident dependent children, where at least two are between the ages 5-14 years and any subsequent children are greater than 4 years	Dummy
ch G	Two or more resident dependent children, where one is between the ages 5-14 and any subsequent children are greater than 14 years	Dummy
ch H	Two or more resident dependent children, where all are aged greater than 14 years	Dummy
ch I (base)	No resident dependent children	Dummy
non-resch	Any non-resident children	Dummy
pnon-resch	Partner has any non-resident children	Dummy
wage	Real hourly wage of female (AUD)*	Continuous
wage2	Real hourly wage of female squared (AUD)*	Continuous
pwage	Real hourly wage of male partner (AUD)*	Continuous
non-lbinc	Real non-labour income^	Continuous
rural	Household located in a rural area#	Dummy
gh	Physical Health (from the SF-36)	Index [0:100]
mh	Mental Health (from the SF-36)	Index [0:100]
immi A	Time lived in Australia is 0-4 years	Dummy
immi B	Time lived in Australia is 5-9 years	Dummy
immi C	Time lived in Australia is 10-19 years	Dummy
immi D	Time lived in Australia is 20+ years	Dummy
immi E (base)	Australian born	Dummy
mtleave	Paid maternity leave	Dummy
unmtleave	Unpaid maternity leave	Dummy
pptleave	Paternity leave of male partner (paid or unpaid)	Dummy
union	Trade union membership%	Dummy
sector	Employed in the private sector	Dummy
unemprt	Unemployment rate (%) @	Continuous
married	Legally married	Dummy
ind A	Agriculture, Forestry and Fishing Industry†	Dummy
ind B	Mining Industry †	Dummy
ind C	Manufacturing Industry †	Dummy
ind D	Electricity, Gas and Water Supply Industry †	Dummy
ind E	Construction Industry †	Dummy
ind F	Wholesale Trade Industry †	Dummy

Variable Name	Description	Variables Required
ind G	Retail Trade Industry †	Dummy
ind H	Accommodation, Cafes and Restaurants Industry †	Dummy
ind I	Transport and Storage Industry †	Dummy
ind J	Communication Services Industry †	Dummy
ind K	Finance and Insurance Industry †	Dummy
ind L	Property and Business Services Industry †	Dummy
ind M	Government Administration and Defence Industry †	Dummy
ind N	Education Industry †	Dummy
ind O	Health and Community Services Industry †	Dummy
ind P	Cultural and Recreational Services Industry †	Dummy
ind Q (Base)	Personal and Other Services Industry †	Dummy
NSW	New South Wales	Dummy
VIC	Victoria	Dummy
QLD	Queensland	Dummy
SA	South Australia	Dummy
WA	Western Australia	Dummy
TAS	Tasmania	Dummy
NT	Northern Territory	Dummy
ACT (base)	Australian Capital Territory	Dummy

Note: (1) Dummy variables are coded so that presence is set to one and absence to zero. (2) Index [0:100] is an index measured as a continuous variable with range 0 to 100. (3) * The hourly wage rate is inflated to the value in the year 2006 by the RBA annual inflation rate over the period (2001-2006). (4) ^ Non-labour income is inflated to the value in the year 2006 by the RBA annual inflation rate over the period (2001-2006). (5) # Rural location of a household is defined by the ABS Australian Standard Geographical Classification (2001), Cat. No. 1216.0, based on population counts of Census Collection Districts (CD). (6) % Trade Union membership as defined by the ABS. (7) @ The unemployment rate is derived from Data Cube LM8–Labour Force Status by Sex, State, Age, Marital Status (ABS Labour Force, Australia, Detailed – Electronic Delivery, Mar 2008, Cat. No. 6291.0.55.001). (8) † Industry classifications are defined by the ABS Australian and N. Z. Standard Industrial Classification (ANZSIC) 1-digit code, first edition (1994), Cat. No. 1293.0.

Table 3: Summary Statistics (Australia)

	Couple Females				Single Females			
	Employment		**Hours Supplied**		**Employment**		**Hours Supplied**	
	Mean	*Std. Dev.*	Mean	*Std. Dev.*	Mean	*Std. Dev.*	Mean	*Std. Dev*
empt	0.547	*0.498*	-	-	0.598	*0.490*	-	-
lnhours	-	-	3.336	*0.563*	-	-	3.448	*0.549*
empt_lag	0.574	*0.495*	-	-	0.624	*0.484*	-	-
exp	16.878	*10.266*	-	-	15.450	*11.842*	-	-
exp2	390.253	*412.863*	-	-	378.900	*456.144*	-	-
jbsearch	0.424	*1.488*	-	-	0.761	*2.137*	-	-
jbsearch2	2.395	*25.447*	-	-	5.144	*32.143*	-	-
age 18-24 (base)	0.047	*0.212*	0.049	*0.216*	0.184	*0.387*	0.209	*0.407*
age 25-34	0.238	*0.426*	0.260	*0.439*	0.208	*0.406*	0.217	*0.412*
age 35-44	0.299	*0.458*	0.353	*0.478*	0.210	*0.408*	0.224	*0.417*
age 45-54	0.217	*0.412*	0.259	*0.438*	0.223	*0.417*	0.244	*0.430*
age 55-64	0.199	*0.399*	0.080	*0.271*	0.175	*0.380*	0.105	*0.307*
page 18-24 (base)	0.029	*0.168*	0.031	*0.174*	-	-	-	-
page 25-34	0.197	*0.397*	0.223	*0.416*	-	-	-	-
page 35-44	0.286	*0.452*	0.320	*0.467*	-	-	-	-
page 45-54	0.227	*0.419*	0.288	*0.453*	-	-	-	-
page 55+	0.261	*0.439*	0.138	*0.345*	-	-	-	-
ed 1	0.260	*0.438*	0.365	*0.482*	0.243	*0.429*	0.338	*0.473*
ed 2	0.092	*0.290*	0.102	*0.302*	0.092	*0.289*	0.114	*0.318*
ed 3	0.114	*0.318*	0.119	*0.324*	0.134	*0.341*	0.143	*0.350*
ed 4	0.173	*0.378*	0.171	*0.377*	0.189	*0.391*	0.195	*0.396*
ed 5 (base)	0.361	*0.480*	0.243	*0.429*	0.342	*0.475*	0.211	*0.408*
ped 1	-	-	0.321	*0.467*	-	-	-	-
ped 2	-	-	0.112	*0.315*	-	-	-	-
ped 3	-	-	0.275	*0.446*	-	-	-	-
ped 4	-	-	0.110	*0.313*	-	-	-	-
ped 5 (base)	-	-	0.183	*0.386*	-	-	-	-
ch A	0.080	*0.272*	0.077	*0.266*	0.043	*0.202*	0.021	*0.144*
ch B	0.039	*0.194*	0.050	*0.218*	0.066	*0.248*	0.064	*0.245*
ch C	0.057	*0.231*	0.063	*0.243*	0.065	*0.246*	0.073	*0.260*
ch D	0.078	*0.268*	0.049	*0.217*	0.018	*0.133*	0.005	*0.068*
ch E	0.085	*0.279*	0.075	*0.263*	0.040	*0.196*	0.021	*0.144*

	Couple Females				Single Females			
	Employment		Hours Supplied		Employment		Hours Supplied	
	Mean	*Std. Dev.*	Mean	*Std. Dev.*	Mean	*Std. Dev.*	Mean	*Std. Dev*
ch F	0.150	*0.357*	0.180	*0.384*	0.073	*0.260*	0.061	*0.240*
ch G	0.052	*0.223*	0.068	*0.251*	0.035	*0.185*	0.034	*0.181*
ch H	0.046	*0.209*	0.064	*0.246*	0.024	*0.152*	0.027	*0.161*
ch I (base)	0.413	*0.492*	0.374	*0.484*	0.637	*0.481*	0.693	*0.461*
non-resch	0.342	*0.475*	0.235	*0.424*	0.325	*0.469*	0.242	*0.429*
pnon-resch	0.387	*0.520*	0.295	*0.456*	-	-	-	-
non-lbinc	123.984	*237.857*	68.279	*220.727*	196.528	*242.936*	109.162	*211.496*
wage	-	-	23.469	*10.543*	-	-	21.420	*9.218*
wage2	-	-	661.930	*778.906*	-	-	543.750	*624.398*
pwage	20.343	*16.666*	24.344	*14.669*	-	-	-	-
rural	0.163	*0.369*	0.142	*0.349*	0.092	*0.289*	0.074	*0.262*
gh	70.976	*21.220*	75.077	*18.043*	66.745	*22.785*	72.229	*19.215*
mh	74.266	*16.735*	76.086	*14.847*	69.112	*19.383*	72.714	*16.858*
immi A	0.016	*0.127*	0.014	*0.117*	0.010	*0.102*	0.009	*0.095*
immi B	0.024	*0.154*	0.023	*0.150*	0.014	*0.118*	0.012	*0.110*
immi C	0.055	*0.228*	0.056	*0.229*	0.039	*0.194*	0.038	*0.191*
immi D	0.134	*0.341*	0.115	*0.319*	0.114	*0.318*	0.098	*0.298*
immi E (base)	0.770	*0.421*	0.793	*0.405*	0.822	*0.383*	0.842	*0.365*
mtleave	-	-	0.483	*0.500*	-	-	0.502	*0.500*
unmtleave	-	-	0.740	*0.439*	-	-	0.710	*0.454*
pptleave	-	-	0.648	*0.478*	-	-	-	-
unemprt	3.304	*1.811*	-	-	7.818	*3.002*	-	-
union	-	-	0.333	*0.471*	-	-	0.324	*0.468*
sector	-	-	0.611	*0.488*	-	-	0.643	*0.479*
married	0.832	*0.374*	0.815	*0.388*	-	-	-	-
ind A	-	-	0.011	*0.102*	-	-	0.007	*0.086*
ind B	-	-	0.001	*0.037*	-	-	0.004	*0.066*
ind C	-	-	0.050	*0.219*	-	-	0.064	*0.244*
ind D	-	-	0.003	*0.054*	-	-	0.003	*0.056*
ind E	-	-	0.015	*0.120*	-	-	0.010	*0.101*
ind F	-	-	0.026	*0.159*	-	-	0.022	*0.146*
ind G	-	-	0.096	*0.294*	-	-	0.119	*0.323*
ind H	-	-	0.037	*0.189*	-	-	0.063	*0.243*
ind I	-	-	0.021	*0.145*	-	-	0.016	*0.124*

	Couple Females				Single Females			
	Employment		Hours Supplied		Employment		Hours Supplied	
	Mean	*Std. Dev.*	Mean	*Std. Dev.*	Mean	*Std. Dev.*	Mean	*Std. Dev*
ind J	-	-	0.020	*0.141*	-	-	0.019	*0.136*
ind K	-	-	0.052	*0.221*	-	-	0.047	*0.211*
ind L	-	-	0.105	*0.307*	-	-	0.095	*0.294*
ind M	-	-	0.063	*0.243*	-	-	0.068	*0.251*
ind N	-	-	0.215	*0.411*	-	-	0.161	*0.368*
ind O	-	-	0.229	*0.420*	-	-	0.235	*0.424*
ind P	-	-	0.025	*0.156*	-	-	0.027	*0.161*
ind Q (base)	-	-	0.031	*0.174*	-	-	0.041	*0.198*
NSW	0.293	*0.455*	0.303	*0.460*	0.295	*0.456*	0.286	*0.452*
VIC	0.239	*0.426*	0.250	*0.433*	0.238	*0.426*	0.254	*0.436*
QLD	0.212	*0.409*	0.200	*0.400*	0.217	*0.412*	0.211	*0.408*
SA	0.090	*0.286*	0.085	*0.279*	0.102	*0.303*	0.087	*0.282*
WA	0.099	*0.298*	0.083	*0.277*	0.086	*0.281*	0.090	*0.286*
TAS	0.033	*0.179*	0.031	*0.172*	0.035	*0.184*	0.040	*0.195*
NT	0.007	*0.085*	0.012	*0.109*	0.009	*0.092*	0.011	*0.103*
ACT (base)	0.027	*0.163*	0.035	*0.184*	0.018	*0.131*	0.022	*0.146*

Notes: (1) Means are for the pooled data (i.e. six waves 2001-2006). (2) *Std Dev* represents the standard deviation.

Although many of the independent explanatory variables (i.e. *x*) included in the analysis for this Paper are common in previous labour supply models (e.g. level of education, marital status, and wage) there are a number that have, generally, not been included in previous work, or they are defined to a greater level of detail in this Paper:[17]

- In the *Hours* equation for couple households, separate dummy variables representing maternity leave are included; specifically:[18]
  - female's paid maternity leave
  - female's unpaid maternity leave
  - male partner's paternity leave
- A dummy variable is included to represent union membership.
- In contrast with many other studies, in this analysis the children are represented by sets of dummy variables (not a count), and therefore do not assume a linear relationship. Dummy variables cover:
  - children 0 to 4 years
  - children 5 to 14 years
  - children 15 to 24 years.
- For the equations, dummy variables differentiating between one and two or more dependent children residing in the household:
  - one child 0 to 4
  - two or more children, where at least the first two are 0 to 4 and any subsequent children are 0+
  - two or more children, where one is 0 to 4 and any subsequent children are 5+
  - one child 5 to 14
  - two or more children, where at least the first two are 5 to 14 and any subsequent children are 5+
  - two or more children, where one is 5 to 14 and any subsequent children are 15+
  - one child 15 to 24

[17] Noting that with all studies based on survey data, a potential drawback to increasing the number of explanatory variables is the increase in prevalence of missing data and hence reduced the sample size.

[18] Note that maternity leave dummy variables (and health variables below) are taken from the HILDA self-completion survey and are responsible for a reduction in sample size.

- two or more children, where all are aged 15+

- A dummy variable for non-resident (own and partner's) children was included.

- In place of the usual dichotomous dummy variable (zero if not an immigrant, one if an immigrant), in this analysis immigrants are represented by a set of dummy variables, which do not assume a linear relationship[19]:
  - lived in Australian for 0 to 4 years
  - lived in Australian for 5 to 9 years
  - lived in Australian for 10 to 19 years
  - lived in Australian for 20+ years

- Health variables are taken from the questions for the Short Form (36) Health Survey (SF-36).[20] They provide continuous measures (scales range between zero and 100). There is discussion in the literature regarding the appropriateness of including health as it may be endogenous The relationship between the hours supplied decision and health may be endogenous (i.e. the direction of causality is unclear—poor health may reduce hours, or an involuntary reduction in hours may cause distress or poor health), but there was little evidence of endogeneity in estimated models. Health index variables are included for:
  - mental health
  - physical health

- In couple female specifications, other measures for the partner's attributes included are:
  - education
  - employment status
  - age

- Industry sector is represented by dummy variables.

- State/Territory dummy variables are included.

- A dummy for rural versus urban living is included.

---

[19] English language ability was also considered for inclusion in the models but small 'cell' numbers (i.e. a very unbalanced distribution between 0 and 1 for the dummy variable) caused the software to exclude the variable.

[20] The SF-36 consists of two summary measures calculated from eight scale scores (physical functioning, role physical, bodily pain, general health perceptions, vitality, social functioning, role emotional, and mental health). The summary measures, or scales, are the physical component score and the mental component score (see http://www.sf-36.org for further details).

- Non-labour income is included.
- A dummy for public-private sector employment is included.
- Total time out of employment (representing de-skilling and strength of attachment to the labour force) is included.
- Total years of labour force experience are included.
- A state specific (for age group, gender, and marital status) measure of the unemployment rate is included to capture macroeconomic conditions (Wachter (1974)).

In a number of cases the inclusion of the variables noted above is important (e.g. the availability of maternity or paternity leave is generally statistically significant). In some cases, however, explanatory variables are not statistically significant, but in several cases their inclusion adds to the understanding of labour supply (by suggesting that manipulating that individual attribute will not influence hours of work supplied). For example, being a rural resident does not influence the probability of labour force participation suggesting that extra services in the rural sector to encourage participation are not warranted. On the other hand, not surprisingly, the unemployment rate was found to be non-significant, probably because, during the 6 years of the HILDA data, unemployment had been low by historical standards, and thus the model should maintain this variable in the face of future higher rates of unemployment.

Econometric Model Results

The section below presents and discusses the results of the labour supply model (results are, for ease of access, restricted to coefficients and an indicator of statistical significance.

The results are reported for single and coupled females in Australia—as the results demonstrate, the labour market behaviour of single females is different to those who reside with a male partner.[21] Single females are more likely to be labour force participants and are more likely to be employed than partnered females (i.e. they are more likely to allocate their time to work due to their limited family responsibilities to a male partner/husband and to children) and their limited access to alternative sources of income (e.g. a partner's income). The behaviour of single females is similar to that of single males, whereas the behaviour of partnered females differs importantly to that of partnered males.

Employment Equation

The labour force participation model examines the impact of selected explanatory (independent), and control variables, on females' probability of labour force participation: where participation is defined as those who are employed, relative to the total relevant population of females.

The fitted parameter estimates, for single and couple females, are reported in Table 4 below. As the results are estimates of a random effects limited dependent variable (probit) panel data regression model, interpreting the estimated coefficients is not straightforward due to the non-linear nature of the underlying (probit) distribution function. Consequently, the parameter estimates are reported as the conversion to the marginal effects,[22] calculated at the variables' sample means.[23] For the continuous variables, including index variables, the coefficients are interpreted as the effect on the probability of labour force participation for a small (marginal) change in an explanatory variable. For discrete (dummy) variables, the coefficients are interpreted as the effect on the probability of labour force participation for a change from one state to the other (i.e. between zero and one) for that dummy explanatory variable.[24]

---

[21] Same sex couples are a very small proportion of the HILDA sample and are excluded from this Paper on econometric grounds.

[22] More completely, from the probit model, the values presented in Tables 4 are defined as $\partial \Pr[\text{Participation}_{it} = 1 \mid x_{it}]/\partial x_{it} = F'(x'_{it}\beta)\beta_j$ (i.e. the partial derivative of Participation with respect to an individual explanatory variable, x). This specification demonstrates that the marginal effect differs depending on the value of the explanatory variables. Note that the ratio of coefficients is equal to the ratio of the marginal effects (i.e. the ratio of marginal effects is equal to the relative effects of changes in repressors) (Cameron and Trivendi 2005).

[23] More specifically the marginal effect is the change in the conditional mean of the probability of labour force participation when the explanatory variable (continuous, index, and dummy) changes by one unit.

[24] Note that work-related variables are not included in the participation equation as they are missing (i.e. not relevant) for non-participants and the unemployed (they cannot be included by assigning zeros to the missing values as this causes a spurious statistical relationship between those characteristics and participation).

Results—Employment Equation

Table 4 provides model details for participation equations for single and couple females.

Table 4: Employed Equation, Single and Couple Females—Australia

	Couple Females		Single Females	
empt_lag	1.791	***	1.763	***
	(0.045)		*(0.061)*	
wave A	-		-	
	-		-	
wave B	-0.088		0.033	
	(0.064)		*(0.084)*	
wave C	-0.103		0.051	
	(0.064)		*(0.084)*	
wave D	0.007		-0.006	
	(0.065)		*(0.087)*	
wave E	0.134	**	0.186	**
	(0.066)		*(0.089)*	
wave F	0.110	*	0.204	**
	(0.067)		*(0.092)*	
exp	0.088	***	0.101	***
	(0.010)		*(0.013)*	
exp_sq	-0.001	***	-0.002	***
	(0.000)		*(0.000)*	
jbsearch	-0.067	**	-0.105	***
	(0.024)		*(0.028)*	
jbsearch_sq	0.003	**	0.004	**
	(0.001)		*(0.002)*	
age 18-24	-		-	
	-		-	
age 25-34	-0.021		-0.160	
	(0.137)		*(0.107)*	
age 35-44	-0.307	*	-0.585	***
	(0.161)		*(0.145)*	
age 45-54	-0.653	***	-0.966	***
	(0.187)		*(0.174)*	
age 55-64	-1.271	***	-1.426	***
	(0.211)		*(0.203)*	

	Couple Females		Single Females	
page 18-24	-		-	
	-		-	
page 25-34	-0.184		-	
	(0.145)		-	
page 35-44	-0.299	*	-	
	(0.159)		-	
page 45-54	-0.254		-	
	(0.172)		-	
page 55+	-0.695	***	-	
	(0.188)		-	
ed 1	0.529	***	0.592	***
	(0.062)		*(0.087)*	
ed 2	0.241	**	0.346	**
	(0.075)		*(0.105)*	
ed 3	0.258	***	0.255	**
	(0.070)		*(0.087)*	
ed 4	0.208	***	0.262	**
	(0.060)		*(0.080)*	
ed 5	-		-	
	-		-	
ch A	-0.720	***	-0.403	**
	(0.083)		*(0.134)*	
ch B	-0.027		0.161	
	(0.115)		*(0.115)*	
ch C	0.227	**	0.507	***
	(0.098)		*(0.127)*	
ch D	-0.859	***	-0.123	
	(0.094)		*(0.224)*	
ch E	-0.370	***	0.165	
	(0.089)		*(0.150)*	
ch F	-0.017		0.252	**
	(0.080)		*(0.120)*	
ch G	0.301	**	0.334	**
	(0.111)		*(0.158)*	
ch H	0.261	**	0.345	*
	(0.114)		*(0.191)*	

	Couple Females		Single Females	
ch I	-		-	
	-		-	
non-resch	-0.164	**	-0.083	
	(0.076)		*(0.081)*	
pnon-resch	0.107	*	-	
	(0.057)		-	
non-lbinc	-0.001	***	-0.002	***
	(0.000)		*(0.000)*	
pwage	0.003	**	-	
	(0.001)		-	
rural	-0.068		-0.057	
	(0.057)		*(0.091)*	
gh	0.006	***	0.008	***
	(0.001)		*(0.001)*	
mh	0.002	*	0.006	***
	(0.001)		*(0.002)*	
immi A	-0.392	**	-0.698	**
	(0.147)		*(0.222)*	
immi B	-0.086		-0.551	**
	(0.129)		*(0.225)*	
immi C	-0.256	**	-0.181	
	(0.090)		*(0.132)*	
immi D	-0.116	*	-0.180	**
	(0.066)		*(0.091)*	
immi E	-		-	
	-		-	
unemprt	-0.008		0.008	
	(0.015)		*(0.012)*	
married	-0.101	*	-	
	(0.060)		-	
Intercept	-1.437	***	-1.868	***
	(0.233)		*(0.277)*	
Sample	9364		5820	
Individuals	3497		2332	
pseudo-R^2	0.349		0.380	
Log Likelihood	-2961.734		-1676.781	

Notes: (1) See Table 2 above for variable name legend. (2) Separate dummy variables are included for 1 or 2 children. (3) *** represents *p*-value ≤ 1%, ** represents *p*-value ≤ 5% (> 1%), * represents *p*-value ≤ 10% (> 5%). (4) Data are marginal effects for probit model (see text). (5) Sample is the number of observations (i.e. spread across 6 waves of the HILDA). (6) Control variables for state are excluded.

Dynamics and State Dependency (Lagged Labour Force Participation): As is common in most labour force participation econometric models for Australia (and similar OECD countries), current labour force participation and employment is influenced strongly by previous labour force status—the lagged value of employment status (*empt_lag*) is statistically significant at better than the 0.001% level (p-value 0.000). The inclusion of the lagged dependent variable indicates the presence of *state dependence* as the other parameters effects on participation partly operate through lagged participation. Omitting this variable from the model specification generally increases the apparent statistical significance of other explanatory variables and increases the size of the estimated coefficients, moreover econometric model results are biased and not reliable (this is a missing variable model misspecification).

Trend: There is little evidence of a trend in employment over the six-year period when other factors are controlled (i.e. any actual trend is accounted for by other explanatory variables). Only four of 10 wave dummy (wave A to wave F) variables are statistically significant.

Labour Market Experience: In all specifications, the variables representing years of work experience (*exp*) and years of experience squared (exp^2), intended to capture backward bending supply due to decreasing returns to years of work experience on participation, are statistically significant. These results are common in participation and employment equations. For both single and couple females, the impact of experience is much larger than the backward bending impact, but the backward bending impact does not occur until the latter part of working life—that is, 32 years for single females, and 40 years for couple females. Thus, each additional year of experience increases the probability of participation at a decreasing rate up to about 32 years, after which time the impact of the backward bending effect dominates and additional work experience reduces the probability of participation.

For the probit models for employment of females, the 'backward bending' effect is probably immaterial and it is more useful to focus on the impact of *exp* with coefficients ranging from 0.09 to 0.1 (or 9% to 10%). When considering the 95% confidence interval for the point estimates for the coefficient, there is no statistical difference of the impact of *exp* on single females or coupled females.

Seeking Employment: In addition to labour force experience, the impact of duration of job search on the probability of participation behaves accordingly in an inverse fashion (*jbsearch*, and the square of the period seeking employment, $jbsearch^2$ – intended to capture the increasing cost of unemployment on participation). In both specifications, *jbsearch* and $jbsearch^2$ are statistically significant and have a strong 'forward bending' influence. As expected, the length of time spent unemployed negatively affects the probability of gaining employment, 7% for couple females and 11% for single females. However, this quickly diminishes to zero within 11 to 12 years, at which point it begins to have a small increasing influence.

Education: The influence of increased levels of educational attainment on the probability of employment is both statistically and economically significant for both couple and single females. As expected, the positive influence of higher education levels of education attainment increases at an increasing rate, above the base education level (i.e. Year 11 and below). Interestingly, the coefficient magnitudes and statistical significances are similar for both couple and single females. Level of education attainment positively influences the probability of employment by 0.21 to 0.26 (21% to 26%) at the Year 12 and Certificate I/II qualification level, through to 0.53 to 0.59 (53% to 59%) for university qualifications (for couple and single females, respectively). Note, however, that the relationship between education and participation is not strictly linear, with a large gap between the influence of the vocational Advanced Diploma/Diploma qualification level and the university obtained qualification levels. In general, the model results show that for couple females, education and increasing levels of education dramatically increase participation despite the presence of a male partner (with or without children in the household).

Children at Home: The influence of children on the labour market participation of couple and single females is interesting. Having only one child below the age of 5 years (*ch A*) severely reduces the probability of participation of both couple-and single females—that is, 0.72 (72%) for couple females (at the 1% level of statistical significance) and 0.43 (43%) for single females (at the 5% level of significance). However, for single females, the presence of more young children does not statistically influence the probability of employment. On the other hand, couple females are more inclined to stay out of the workforce when they have two or more children and at least one is less than 5 years of age (*ch D* and *ch E*)—reduces the probability of employment between the range 0.86 and 0.37 (86% and 37%). For both couple and single females, the presence of older child/children (aged between 5-14 and 15-24 years) increases the probability of employment, particularly for single females. For example, one child aged 15-24 years (*ch C*) increases the probability of employment by 0.23 to 0.51 (23% to 51%) for couple and single females (statistically significant at the 5% and 1% levels, respectively); having two or more children aged over 14 years increases the probability 0.26 to 0.35 (26% to 35%) (statistically significant at the 5% and 10% levels, respectively). In summary, the results indicate that couple females, relative to single females, exit the labour force to have children and to care for their very young, returning to the labour force when children attend school or higher education. This could possibly be a consequence of the male "breadwinner" effect. Generally, very young children reduce the probability of female participation in the labour force, but older children encourage participation—perhaps, for example, because of the cost of raising children, the desire to resume a profession, or for social contact.

Non-residential Children: the non-resident children, of any age, of couple females and their partners (*non-resch* and *pnon-resch*) have a significant impact on couple female participation as expected. The presence of coupled female non-resident children (*non-resch*) decreases the probability of their participation by 16%; whereas the non-resident children of couple females' partners (*pnon-resch*) increase the participation of couple females by 11%.

Non-labour Income: As is commonly suggested, access to non-labour real income per week (*nonlbinc*) obtained from investment income; private and public transfer income; and private and foreign pension income reduces the probability of females participating in the labour force. This variable is statistically significant at better than the 0.01% level in both specifications. Although the estimated coefficients are small (ranging between -0.001 and -0.002), when translated to the impact due to a $100 increase in non-labour income per week, the impact is a 10% to 20% reduction in the probability of participation. Clearly, this is a control variable as there is no, general, desire to reduce non-labour income and hence no acceptable policy objective—but the impact is large and hence it is important to note.

Immigrants' Residential Period: Although the years of residence of immigrants are not itself subject to policy intervention, factors relating to the impact of length of residence can be influenced. The literature suggests that the length of residence of immigrants may proxy a number of other attributes. Some attributes are measurable (e.g. English language ability), some not (e.g. entrepreneurial attitude), and some are subject to influence (e.g. English ability, knowledge of Australian institutions and knowledge of the labour market processes). Thus, immigrants' labour market status represents their adjustment to the Australian labour market (Chiswick *et al.* (2005)) (see Lester (2008) for a review).

In both specifications, the set of immigration dummy variables (*immi A – D*) are both statistically and economically significant, relative to the base (i.e. born in Australia). The influence of the immigration dummies on employment highlight the non-linear nature of the length of time spent living in a foreign country and the imperative of successful engagement in the labour market. For newly migrated females (*immi A*), the probability of employment is initially low—that is, 39% for couple females and 70% for single females. However, for female immigrants that had lived in Australia for 20 years or more (*immi D*), their probability of employment, while still negative, drastically improved—that is 12% for couple females and 18% for single females. Interestingly, after a 20 year duration living in Australia, the probability of employment of both couple and single females were at similar levels, although still negative. Furthermore, the rate at which the probability of employment increased as the length of time single females lived in Australia increased is impressive, relative to couple females. Again, this could possibly be a consequence of the male "breadwinner" effect.

Thus, while the years of residence of immigrants cannot be influenced, government-provided access to English language tuition, job search knowledge including information about the operation of the Australian labour market, and other social capital formation may increase the probability of participation of immigrant females to that of otherwise similar non-immigrants. This is an area where further research may be valuable.

Health: Two measures of health are included in the models, general physical health (*gh*) and mental health (*mh*) as indexes with a range [0:100].

Both measures of health are statistically significant for both couple and single females. However, the magnitudes of the coefficients and their influence on the probability of employment are not large. The estimated coefficients of general health (*gh*) imply that for a unit increase in the index (i.e. improving health) increases the probability of participation by 0.6%, for couple females, and 0.8%, for single females. Similarly, mental health (*mh*) positively influences the probability of participation by 0.2% and 0.6%, for couple and single females, respectively.

Control variables

In the context of the Employed equation, control or covariate variables are included to control for known (or expected) influential characteristics or factors that, if excluded, bias econometric estimates; however, there is no scope to influence them, and hence beyond policy control consideration (although in some case, their impact is interesting).

Females' Partner's Attributes: For couple-females, inclusion of partner attributes (such as education, wage or salary, and non-resident children) are control variables (variables with a *p* prefix, e.g. *pwage*). Interestingly, the partner's attributes appear to have less impact than expected.

As previously discussed, the dummy variable representing the partner's non-resident children (*pnon-resch*) has a minor positive influence on the couple female's probability of employment; whereas the influence of non-resident children, of the couple females, (*non-resch*) has a minor negative impact on couple females probability of employment. These estimates may be capturing the effects of unconventional family compositions (i.e. the increased prevalence of split familles and divorce) and how the caring responsibilities are allocated.

The inclusion of the set of dummy variables capturing the effect of the partner's age (*page 25-34 – 55+*), on the couple females' probability of employment, appears to only be significant negative influence once the male partner 55 years or more—that is, 70% for couple females.

Interestingly, the male partners hourly wage rate (*pwage*) has a slight positive influence on the probability of employment of couple females—that is, for a $1 increase in the partner's hourly wage, the participation of couple females increases by 0.3%.

Finally, the influence of being legally married (*married*) is statistically significant, at the 10% level, and has a negative affect on the probability of participation. For the sample of couple females, the impact is a 10% reduction.

The tendency for inter-dependence of female labour force participation and a male resident partner or spouse indicates that further research using "collective" labour supply models to obtain more efficient and robust estimates, and to observe intra-household welfare allocations, is appropriate—when the limitations imposed by currently available theory and software can be overcome.

Single females versus couple females—a summary

It is useful, as a final step, for examination of the influence on the probability of female labour force participation, to summarise and compare the model estimates for single and couple females at a more general level. Of interest is whether, as is conventional wisdom, there is empirical evidence that single and couple females have different patterns of labour force participation. It is clear from the models that there are surprising similarities.

Similarities for Australian single and couple females are:

- The control for state dependency (*empt_lag*) is necessary to correctly model single and partnered females (the absence of this control cause the importance of explanatory variables to be overstated). This control has a similar impact for single and couple females (that is, the 95% confidence intervals for single females and couple females coincide indicating no statistical difference in the estimated coefficients at the 1% level of significance[25]).

- Trends (via the *wave* dummies) have little influence.

- Years of labour market experience (*exp*) (and experience squared, exp^2) impacts do not differ substantially (the 95% confidence intervals for single females and couple females coincide).

- Similar to labour market experience, the period of job search (*jbsearch* and $jbsearch^2$) impacts do not differ substantially.

- Education (*ed 1 – ed 4*) matters, the impact of each level of education are comparable, and there is little difference between single and couple females (the 95% confidence intervals coincide).

- There is little difference in the impact of general physical health (*gh*) and mental health (*mh*) (the 95% confidence intervals coincide).

Differences for single and couple females are:

- There are important differences in the impact of children—which is not itself a factor that can be manipulated by policy to any great extent (particularly in the short-run), but indirectly the impact of children can be influenced by, for example, the provision of childcare:

    - Two or more resident dependent children where at least two a between 0 to 4 years of age and any subsequent children are greater than 0 years (*ch D*), reduce participation of couple females, but do not appear to alter the behaviour of single females.

    - Two or more resident dependent children where one is between 0 to 4 years of age and any subsequent children are greater than 4 years (*ch E*), reduce the participation of couple females, but do not appear to alter the behaviour of single females.

    - Two or more resident dependent children where at least two are between 5 to 14 years of age and any subsequent children are greater than 4 years (*ch F*), increase the participation of single females, but do not appear to alter the behaviour of couple females.

    - Two or more resident dependent children where at least two are between 5 to 14 years of age and any subsequent children are greater than 14 years (*ch G*), and two or more resident dependent children where all are greater than 14 years (*ch H*) have a similar, positive, impact on the participation of both single and couple females (the 95% confidence intervals coincide).

---

[25] The 95% confidence interval for the estimated coefficient, β, is constructed as $\beta \pm$ (standard error of β) * (*z*-value for 95%=1.96).

- Access to non-labour income (*non-lbinc*) has a significantly larger impact for single females compared to couple females (e.g., a $100 increase reduces participation for single females by 20%, but by only 10% for couple females).

In summary, although there are a number of similarities in the model estimates for single females and couple females, there are sufficient differences to confirm that failure to model singles and couples separately is an aggregation problem which results in *aggregation bias* (Greene 2003)—leading to potentially incorrect inference and misguided policy analysis and recommendations.

Policy implications arising from the analysis of female labour force participation tend to follow the literature—there are limits to potential intervention, and most policy can at best be directed to longer-term issues. For example, generally, education increases the probability of labour force participation; However, education (and associated vocational skills development) is not subject to short-run manipulation. Similarly, very young children in a household reduce the participation of females, but whether there is a long-term advantage to pursue methods to increase the participation of this group is a complex question, as is the issue of what influences the decision to have a child and its relationship to labour market participation.

Finally, a number of factors that influence participation have not been considered in this Paper due either to their being out of scope of this Paper and/or a lack of suitable data. For example, participation is influenced by availability of apprenticeships, access to educational institutions and the range of courses they offer (Richardson & Teese (2006)). In addition, longer-term demographic changes influence participation: for example, projections suggest that there will be little change in the number of young people entering the labour force, but the ageing of Australia's population means more people will retire from the labour force suggesting an increased need for relatively older workers (Tan and Richardson (2006)). More generally, it is clear that there are significant limitations when trying to forecast labour supply and demand, particular at the regional level. Thus, the complexity and uncertainty generally result in complicated large scale, data intensive and costly modelling methods, such as computable general equilibrium models (Tan, Lester *et al.* (2008)), which are beyond the scope of this Paper.[26] In addition, the impact of issues such as the *discouraged worker* effect (Pissarides (1976)) and *hidden unemployment* (which alter the real and Papered participation rate) and *underemployment* (Wooden (1993)) are also beyond the scope of this Paper but may be avenues for further research.

[26] The issues of reservation wage impacts are not addressed in this Paper. There is no reason to expect a significant change in behaviour, and there were no unexpected changes in the wage distribution, during the period of this analysis (2001-2006) which suggests the complexity of computing implied (consistent) estimates of females and partners market wage (which requires a full maximum likelihood approach to correct for selection effects) would improve this analysis.

Hours Supplied Equation

The results of the labour hours supplied (primary) equations (*Hours*, equation [1] above) are the outcome of the panel two-step estimation procedure. As discussed previously, the model examines the impact of selected determinants on females' supply of hours of paid work per week for the sub-sample of females who are employed. Primary equation estimates include a variable derived from the reduced form (secondary) equation—the probability of employment (*Employed*, equation [2] above). The correction terms used in the hours supplied equation are derived from females' probability of employment rather than their probability of labour force participation (which also includes the unemployed) because the participation decision indicates their willingness to work (or more specifically to enter the labour force), but unemployed participants do not supply hours worked or additional hours worked. Hence, the inclusion of the correction terms in the primary hours supplied equations need to account for the selection bias that occurs from estimating the sub-sample of those that Paper positive hours in paid employment, rather than those who intend supplying hours. The derived variables (or *correction terms*) from the *Employed* equation, incorporated in the *Hours* equation, correct for the influence of sample selection (i.e. potential endogeneity). The role of *dynamics* and *state dependency*, are controlled for by the inclusion of a lagged dependent variable in the *Hours* (primary) and *Employed* (secondary) equations, respectively.

The *Hours* supplied equations are estimated by a log-linear[27] ordinary least squares (OLS) regression model. As described previously, the OLS based approach incorporates adjustments from the *Employed* limited dependent variable model designed to account for unobserved heterogeneity and selection bias and is, therefore, equivalent to a panel data estimator.[28] In log-linear models, the parameter estimates (the βs) measure a constant proportional or relative change in hours for a given absolute change in an explanatory variable (i.e. semi-elasticity).[29] Thus, for continuous explanatory variables, when the estimated coefficients are multiplied by 100 the values are interpreted as a percentage change in hours supplied per week, for an additional, or marginal, unit change in the explanatory variable. For discrete (dummy) explanatory variables, the coefficients are interpreted as a percentage change in hours worked for a change in the dummy variable from zero and one (a change in state).

The impact of the individual time-invariant random effect (to deal with individual unobserved heterogeneity), and the time varying effect (to deal with endogeneity and/or selection bias) *correction terms* (see equation [6] above) are statistically significant for the couple female sample (see Table 5 below).

[27] That is, the depended variable is the logarithm of hours of work (applicable as hours worked is greater than zero), the independent or explanatory variables are in levels (or as observed)—such models are also referred to as semi-log models.

[28] That is, the adjustments to the OLS model incorporate the processes included in panel data estimators.

[29] If estimated parameters are large, the impact of the estimated parameters coefficient should be recalculated using exponentials (i.e. percent change in $hours = 100 * [exp(\beta_i \Delta x_i) - 1]$) to avoid an approximation error for the log-linear functions.

Results—Hours Equations

Overall, the estimated specifications appear to be of reasonable fit and have coefficients with the expected signs and magnitudes. The R^2 values indicate reasonable goodness-of-fit in line with the results of other labour supply models from the literature.

As noted above, *Hours* equations are log-linear models and hence coefficient estimates are interpreted as semi-elasticities—that is, the percent change in hours worked for a one-unit change in the explanatory variable or a change from zero to one for a dummy variable (see, e.g. Gujarati (1988) for details).

The *Hours* equations, in Table 5 below, reveal interesting comparisons between single and couple females. For clarity, control variables such as industry sector and States/Territories are excluded from the Table, as are trend (*wave*) and state dependency (*employment*) controls.

Table 5: Hours Equation, Single and Couple Females—Australia

	Couple Females		Single Females	
age 18-24	-		-	
	-		-	
age 25-34	-0.021		0.075	***
	(0.037)		*(0.029)*	
age 35-44	0.006		0.095	***
	(0.043)		*(0.032)*	
age 45-54	-0.004		0.122	***
	(0.047)		*(0.033)*	
age 55-64	-0.070		0.211	***
	(0.059)		*(0.046)*	
page 18-24	-		-	
	-		-	
page 25-34	0.102	**	-	
	(0.042)		-	
page 35-44	0.087	*	-	
	(0.048)		-	
page 45-54	0.076		-	
	(0.051)		-	
page 55+	0.167	**	-	
	(0.058)		-	
ed 1	0.111	***	0.051	
	(0.025)		*(0.033)*	
ed 2	0.059	**	-0.033	
	(0.028)		*(0.033)*	
ed 3	-0.004		0.035	
	(0.025)		*(0.030)*	
ed 4	0.022		-0.032	
	(0.022)		*(0.030)*	
ed 5	-		-	
	-		-	
ped 1	0.010		-	
	(0.022)		-	
ped 2	0.017		-	
	(0.027)		-	
ped 3	0.019		-	

	Couple Females		Single Females	
	(0.020)		-	
ped 4	0.062	**	-	
	(0.026)		-	
ped 5	-		-	
	-		-	
ch A	-0.179	***	-0.163	**
	(0.037)		*(0.083)*	
ch B	-0.151	***	-0.149	***
	(0.033)		*(0.041)*	
ch C	-0.078	**	-0.011	
	(0.031)		*(0.037)*	
ch D	-0.291	***	-0.047	
	(0.048)		*(0.164)*	
ch E	-0.339	***	-0.244	***
	(0.036)		*(0.075)*	
ch F	-0.276	***	-0.208	***
	(0.024)		*(0.048)*	
ch G	-0.220	***	-0.116	**
	(0.031)		*(0.049)*	
ch H	-0.079	**	-0.067	
	(0.030)		*(0.048)*	
ch I	-		-	
	-		-	
non-resch	-0.047	*	-0.022	
	(0.026)		*(0.024)*	
pnon-resch	0.020		-	
	(0.021)		-	
wage	-0.006	**	-0.002	
	(0.003)		*(0.004)*	
$wage^2$	0.000	*	0.000	*
	(0.000)		*(0.000)*	
pwage	-0.003	***	-	
	(0.001)		-	
non-lbinc	0.000		0.000	*
	(0.000)		*(0.000)*	
rural	0.004		0.013	
	(0.020)		*(0.035)*	
gh	-0.001		-0.001	**
	(0.000)		*(0.001)*	
mh	-0.001		-0.001	
	(0.001)		*(0.001)*	
immi A	0.116	**	0.059	
	(0.050)		*(0.088)*	
immi B	0.116	**	0.178	***
	(0.042)		*(0.060)*	
immi C	0.087	**	0.076	*
	(0.028)		*(0.044)*	
immi D	0.033		0.005	
	(0.021)		*(0.027)*	
immi E	-		-	
	-		-	
mtleave	0.147	***	0.100	***
	(0.015)		*(0.018)*	

	Couple Females		Single Females	
unmtleave	0.180	***	0.139	***
	(0.019)		*(0.022)*	
pptleave	-0.037	**	-	
	(0.016)		-	
union	0.120	***	0.079	***
	(0.014)		*(0.018)*	
sector	0.055	**	0.033	
	(0.019)		*(0.022)*	
married	-0.016		-	
	(0.018)		-	
a 1	-0.016	***	0.001	
	(0.004)		*(0.003)*	
a 2	-0.101	**	0.030	
	(0.037)		*(0.026)*	
Intercept	2.571	***	2.859	***
	(0.185)		*(0.212)*	
Observations	5117		3480	
R^2	0.3423		0.3340	

Notes: (1) See Table 2 above for variable name legend. (2) *** represents p-value ≤ 1%, ** represents p-value ≤ 5% (> 1%), * represents p-value ≤ 10% (> 5%). (3) *F*-test is the p-value for the hypothesis test that coefficients are jointly non-significant. (4) *a 1* and *a 2* are the corrections for unobserved heterogeneity and sample selection bias from the *Employed* equations. (5) R^2 (the coefficient of determination) is a measure of goodness-of-fit (1 represents a perfect fit). (6) Sample is the number of individual observations (i.e. spread across 6 waves of the HILDA). (7) Control variables for State/Territory, industry sector, trend, and state dependency are excluded for clarity.

Dynamics and State Dependency (Lagged Employment): Exclusion of (i) the lagged employment variable (*empt_lag*) in the *Employed* equations (Equation [5], above) or (ii) the predicted employment probability forth-order polynomial variables (not listed in Table 5) from the *Hours* equations (Equations [6] and [7], above), caused a noticeable increase in the magnitudes of many of the estimated coefficients in the *Hours* equations. This confirms that failing to account for *dynamics* in the *Employed* equation or *state dependence* in the *Hours* equation, misspecifies the model and biases the estimated coefficients—see Vella and Verbeek (1999).

Trend: As with the *Employed* equations, previously discussed, the set of time (or wave) dummy variables (*wave A – wave E*) are generally non-significant. Nor is there any discernable trend in the hours supplied by couple or single females.

Education: As discussed above, the set of education dummy variables (*ed 1 – ed 4*) have an important influence on the probability of labour force participation. However, it appears that education has little influence on the hours of work of females. Only the two highest levels of education attainment, university qualifications (*ed 4*) and Advanced Diploma/Diploma (*ed 3*) are statistically significant for couple females, and have a small positive effect on hours—that is, an 11% and 6% increase for *ed 4* and *ed 3*, respectively. Thus, no strong pattern of impact of education emerges.

On balance it appears that own and partner's education are not strongly influential in influencing the hours supplied by females, and there appears little if any role for policy intervention.

Children at Home: There is some consistency for results for single and couple females, when considering the impact of own-children living with the female (*ch A*, *ch B*, *ch C*, *ch D*, *ch E*, *ch F*, *ch G*, *ch H*). Thus, for couple-females, the presence of one residential child reduces the hours supplied per week by 16%, for a child aged 0 to 4 years (*ch A*), through to 8%, for child aged 15-24 years (*ch C*). Similarly, for single females, the negative influence on hours supplied per week ranges between 15%, for a child aged 0 to 4 years (*ch A*), through to 14%, for child aged 5-14 years (*ch B*).

Interestingly, the presence of additional resident children and the combination of ages of those children, have non-linear influence on the hours supplied of females. For couple females, the two combinations that include two or more young resident children (*ch D* and *ch E*, see Table 2 for definitions), reduce the hours supplied per week by 25% and 29%, respectively. Whereas, the combination that includes two or more older resident children (*ch H*, i.e. where all are aged greater than 14 years), reduces the hours supplied per week by only 8%. The hours supplied by single females is also influenced in a similar manner, however to a lesser extent than couple females. Additional children in the youngest age group, 0 to 4 years (*ch D*), and oldest age group, 15 to 24 years (*ch H*), are not significant.

Non-residential Children: The presence of non-residential own (*non-resch*) or partner's children (*pnon-resch*) appear to have little impact on hours supplied—that is, for couple-females, *non-resch* decrease hours by about 5%, but are not significant in any other *Hours* equation.

Non-labour Income: Although strongly significant with respect to the probability of employment, non-labour income (*non-lbinc*) is neither statistically nor economically significant with respect to the supply of hours by females. Thus, once non-labour income has influenced females' probability of labour force participation it has no consequential impact on female hours supplied.

Age: The impact of the set of age dummy variables (*age 25-34 – age 55-64*), on hours supplied, are only statistically significant for single females. The estimated coefficients of the four dummy variables, separated at 10 year intervals (i.e. 25-34, 35-44, 45-54 and 55-64 years), behaved in a non-linear fashion increasing the hours supplied at an increasing rate. For example, hours supplied increased for a single female *aged 25-34* by 8%, *aged 35-44* by 10%, *aged 45-54* by 13% and *aged 55-64* by 24% (relative to the base, i.e. *age 18-24*). This is an interesting result that indicates that while getting older decreases females' probability of participation, for those that are employed their engagement with the labour market increases.

Health: Neither measures for general physical health (*gh*) nor mental health (*mh*) had either no statistical or economic significance on hours supplied once considered as an influencing factor in the probability of employment. This may be acceptable if it is thought that those under mental stress self-select out of the labour force, but that explanation is not appropriate for the sample investigated for this Paper. For example, as expected in a measure of the general population, the *mh* score for Australian couple females has a mean of 76 with a minimum of 4 and a maximum of 100 (in the 0 to 100 index). Not all those in stress are self-selected out of the labour force. One possible explanation is that those in mental stress take paid leave and hence do not record a reduction in hours, or maintaining work attachment is seen as part of the treatment for some mental health conditions. While likely to influence some under mental stress this does not appear to be the explanation covering all employed females under mental stress. This is an area that requires further investigation.

Immigrants' Residential Period: As with the *Employed* equations, the set of immigration dummy variables (*immi A – immi D*) are statistically and economically significant. However, the length of duration lived in Australia by immigrant females on hours supplied is statistically more influential for couple females than singles—although the estimated coefficients for both couple and single females are consistent.

In contrast with the *Employed* equations, increasing duration of residence in Australia increases the hours supplied, but at a decreasing rate. For example, couple females that reside in Australia between 0 to 4 years increase their hours supplied by 12%; whereas those that reside between 10 to 19 years increase their hours by 9%. Combining the results of both the *Employed* and *Hours* equations, it can be surmised that female immigrants find it harder to obtain employment initially, relative to Australian born females, and that those that are successful supply more hours. However, as their length of time lived in Australia increases, unobserved barriers that may have limited their engagement with the labour market (i.e. English ability or institutional knowledge etc.) decrease, and their behaviour becomes similar to females born in Australia. For single females this process appears happen at a faster rate than couple females, most likely due to the absence of the support provided by a male spouse or partner.

This is a subject that warrants further investigation.

Wage: The average hours worked by both couple and single females are relatively high, 28 and 31 hours per week, respectively. Consequently, a wage increase may have a limited impact on hours supplied since, Australian families are "time poor" (Apps 2007)—and this is particular so for working mothers. In this case, an increase in wage rates will not necessarily increase hours supplied, and it may result in a reduction in hours worked—i.e. the *backward bending* labour supply curve associated with higher level wage earners. The estimated coefficients reveal that increasing the hourly wage does decrease the hours supplied. However, regardless of the statistical significance of *wage* and $wage^2$, their economic significance on the hours supplied by couple and single females is irrelevant. For couple females, the hourly wage required to significantly reduce their hours is an unrealistic amount. Furthermore, for single females the lack of statistical significance suggests a lack of access to other sources of income curtails their ability to reduce hours, but in parallel with Australia couple females, suggest that they are also "time poor" and choose to maintain hours.

It should also be recalled that in many cases, workers have little control over the number of hours they work—generally, workers have limited discretion on the number of hours worked (even casual employees respond to employers' requests to increase or decrease hours with perhaps little freedom to deny requests).

Moreover, most low-wage workers live in middle and upper income households (Richardson 1998; Harding and Richardson 1999) and hence may choose hours with respect to the household requirements and not based simply on a wage change.[30] As noted

[30] For example, at $10 per hour 15 hours work provides $150 income and an increase to $15 per hour requires only 10 hours for $150—if $150 is sufficient a worker may reduce hours worked.

previously, when the *collective* approach for joint household decisions has been made more accessible, it may be able to provide more empirical insights into the impact of wage on hours supplied.

The issue of the impact of wage on hours supplied appears to be complex than basic theoretical labour economics suggests and requires further investigation to draw conclusions.

Maternity and Paternity Leave: It is clear from the specifications that the availability of paid (*mtleave*) or unpaid maternity leave (*unmtleave*) is an important influence on the hours supplied by females. For both couple and single females the estimated coefficients for both forms of leave are statistically significant at better than the 0.01% level. The presence of paid maternity leave (*mtleave*) increases hours supplied by 16% and 10% for couple and single females, respectively. Furthermore, the influence of unpaid maternity leave on hours supplied is even greater—that is, an increase of 20% and 15% for couple and single females, respectively.

In addition, the paternity leave of couple females' male partners (*pptleave*) is also statistically significant; however, it reduces the hours supplied by couple females by 4%, which is counter-intuitive.

As discussed above, these explanatory variables have rarely been included in previous labour supply models, and thus the consistent statistical significance indicates its absence is a model misspecification (resulting in biased model estimates). Moreover, maternity leave may also be a proxy for other employment conditions (e.g. desirable working conditions).

As maternity leave is an area that could be influenced by government intervention, the importance of the availability of such leave requires further investigation. Thus, for example, as well as more detailed specification of leave entitlements in econometric specifications, the interaction between industry sector and leave could be considered—are there industries where greater attention should be directed?

Control variables

Females' partner's attributes: As with the *Employed* equations, for couple females, inclusion of partner attributes (such as education, wage rate, age, and non-resident children) are control variables. For couple females, the partner's attributes such as wage rate and paternity leave (but not non-residential children, education or age) have an influence on hours. As these are control variables, the estimated coefficients are of limited interest, but the statistical significance of partner's wage in the model makes intuitive sense.

Unlike the *Employed* equation, marital status (*married*) did not statistically influence the hours supplied by couple-females.

Other control variables: There are a number of control variables included in both single female and couple female *Hours* equations. Although there is little if any scope to influence them, directly or indirectly, and hence no avenue for policy intervention, some results are quite interesting. Moreover, where statistically significant, they suggest their absence in previous models is a model misspecification—leading to unreliable econometric results.

For couple and single females, Trade Union membership (*union)* increased hours worked by 13% and 8%, respectively. Furthermore, for couple females, the private sector variable (*sector*) is statistically significant and increased their hours supplied by 6%. The rural location dummy variable (*rural*) did not statistically influence the hours supplied of females.

Single females versus couple females—A summary

As with the analysis of the labour force participation of females, there is empirical evidence that single and couple females have different patterns of labour hours supplied.

Similarities for single and couple females are:

- Trends (via the wave dummies) have little influence (significance shows no pattern for the two groups).
- The influence of additional younger and older children is similar for both couple and single females.
- Non-labour income per week is not economically significant for both couple and single females.
- The impact of general physical health (*gh*) and mental health (*mh*) is not significant for both groups.
- The impact of being an immigrant, as measured by the duration of residence splines, is similar for both groups.
- Maternity leave (*mtleave* and *unmtleave*) impacts are similar.

Differences for single and couple females are:

- The influence of education on hours supplied is only significant for couple females, and only at the higher levels of education attainment (*ed 3* and *ed 4*).
- There is an inverse relationship between wage and hours supplied for couple-females, but not single females—albeit a weak one.

Conclusion

This Paper is based on estimating *Employed* and *Hours* equations for both single and couple females in Australia. The Paper provides justification for the econometric models chosen and discusses the limitations of the models and the ensuing results.

To the extent possible, given current theoretical and applied limitations, this Paper provides models based on recent advances in both theoretical and practical applications of panel (longitudinal) data econometric models. To the extent that the work is an advance on previous methods, it provides econometric model results that are more reliable: it attempts to address possible forms of bias such as model misspecification (including missing variables), unobserved heterogeneity, selection bias, and dynamics and state dependency.

A number of innovations in this Paper (beyond the use of advanced modelling techniques) provide added perspective on the hours supplied decision of females. For example, the results indicate that the availability of maternity leave information is influential on the hours supplied, and the duration of residence in Australia is also influential.

The results clearly indicate that female data must be disaggregated by single and couple-female sub-samples. Although the explanatory power of several important explanatory variables is not different across single and couple female models, a sufficient number differ importantly—it is likely that the aggregation of single and couple females induce aggregation bias and unreliable econometric estimates.

The Paper provides interesting insights to females' behaviour, and suggests several areas where government policy intervention may contribute to increased hours supplied—for example, in the area of maternity leave and access to labour market skills for immigrants. As discussed in the text, advances in theory and econometric practice are likely to provide more sophisticated models (e.g. the *collective* approach) which may lead to further avenues for government intervention.

On the other hand, the probability of labour force participation seems to suggest few areas where state government intervention could successfully influence participation. This is an area that could be considered for further investigation.

Suggestions for further Research

The most important field for future research is to utilise the recent theoretical extension of labour supply modelling, and move beyond the commonly used *unitary* approach to the *collective* decision making modelling method.

As previously discussed, the use of the collective approach observes the decision making process at the individual level, rather than at the household level in the unitary approach. While, not surprisingly, the collective approach has been found not to improve econometric models for single persons, the model consistently, significantly, alters the econometric results for coupled persons. The additional benefit of the collective approach in modelling labour supply is that it takes into account (and in some cases provides methods to extract) the rules or bargaining that takes place within a household (specifically, the intra-household allocations of welfare between male and female partners—which addresses the issue of inequality of decision making power). However, the consequences of the very recent theoretical advances are that a number of impediments to constructing complex collective models exist. For example, extensions of the collective approach to include children and non-labour market participants are still in their infancy and, thus far, do not appear to be fully specified. Moreover, to the extent that collective models have been theoretically solved and hence can be specified for econometric analysis, the estimation of the models require sophisticated computational and econometric techniques beyond those utilised in this Paper, and beyond the more sophisticated 'off-the-shelf' econometric packages. Nonetheless, advanced work is continually appearing in working papers and other sources; and testable specifications—and econometric package add-ons (e.g. STATA ado files)—are expected to become available.

Finally, this Paper has considered econometric models for females, with control for some partner's attributes. An important question—an extension to this Paper—to be considered to further inform the decision making or policy planning process relates to the reaction of male partners to female's changes in participation and hours supplied—if female participation or increased hours was at the expense of a reduction in male participation or hours which sector should be targeted?

Bibliography

ABS (2006a), Australian National Accounts: State Accounts, Catalogue Number 5220.0.

ABS (2006b), Australian Labour Market Statistics, Catalogue Number 6105.0.

ABS (2006c), Australian Labour Market Statistics, Catalogue Number 6105.0.

Apps, P. (2007), 'Taxation and Labour Supply', Centre for Economic Policy Research, ANU, Discussion Paper No. 560.

Aronsson, T., S. Blomquist, and Sacklen, H. (1999), 'Identifying Interdependent Behaviour in an Empirical Model of Labour Supply', *Journal of Applied Econometrics* 14(6): 607-626.

Ashenfelter, O. and Card, D. (eds) (1999), *Handbook of Labor Economics*, Elsevier Science Pub. Co., Amsterdam.

Baltagi, B. H. (2003), *Econometric Analysis of Panel Data*. Chichester, John Wiley & Sons, Ltd.

Barón, J. D. and D. A. Cobb-Clark (2008). 'Occupational Segregation and the Gender Wage Gap in Private- and Public-Sector Employment: A Distributional Analysis', Bonn, IZA Discussion Paper No. 3462.

Bloemen, H. G. (2004). 'An Empirical Model of Collective Household Labour Supply with Nonparticipants', *Discussion Paper* TI 2004-010/3. Amsterdam, Tinbergen Institute.

Blundell, R., Chiappori, P.-A., Magnac, T. and Meghir, C. (2007). 'Collective labour supply: heterogeneity and nonparticipation', *Review of Economic Studies*, vol 74.

Breunig, R., Cobb-Clark, D. A. and Gong, X. (2005), 'Improving the Modeling of Couples' Labour Supply', *IZA Discussion Paper No. 1773*, Bonn.

Butterworth, P., T. Crosier, and Rodgers, B. (2004), 'Mental Health Problems, Disability and Income Support Receipt: A Replication and Extension Using the HILDA Survey', *Australian Journal of Labour Economics* **7**(2): 151-174.

Cameron, A. C. & P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. New York, Cambridge University Press.

Chiappori, P.A. (1988), 'Rational Household Labor Supply', *Econometrica*, Vol. 56, pp. 63-89.

Chiappori, P.A. (1992), 'Collective Labor Supply and Welfare', *Journal of Political Economy*, vol. 100, pp. 437-67.

Chiappori, P.A., and Donni, O. (2005), 'Learning From a Piece of Pie: The Empirical Content of Nash Bargaining', Mimeo, Columbia University.

Chiswick, B. R., Lee, Y. L., and Miller, P. W. (2005), 'A Longitudinal Analysis of Immigrant Occupation Mobility: A Test of the Immigrant Assimilation Hypothesis', *International Migration Review*, 39(2), 332-354.

Couprie, U. (2007), 'Time Allocation Within the Family: Welfare Implications of Life in a Couple', *The Economic Journal*, vol 117 (January): 287-305.

Creedy, J. and G. Kalb (2005), 'Discrete Hours Labour Supply Modelling: Specification, Estimation and Simulation', *Research Paper Number 928*, The University of Melbourne, Department of Economics.

Donnie, O. (2003), 'Collective Household Labor Supply: Nonparticipant and Income Tax', *Journal of Public Economics*, vol 87: 1179-1198.

Donnie, O. (2007), 'Collective Female Labor Supply: Theory and Application', *The Economic Journal*, vol 117 (January): 94-119.

Fortin, B. & G. Lacroix (1997), 'A Test of the Unitary and Collective Models of Household Labor Supply', *The Journal of Political Economy* vol. 107 no. 443, 933-956.

Goza, F. and A. DeMaris (2003), 'Unemployment Transitions among Brazilians in the United States and Canada', *International Migration* **41**(5): 127-150.

Green, S. A. (1991), 'How Many Subjects Does it take to do a Regression Analysis', *Multivariate Behavioral Research* **26**: 499-510.

Greene, W. H. (2003), *Econometric Analysis*, Upper Saddle River, N.J., Prentice Hall.

Gujarati, D. N. (1988), *Basic Econometrics*, New York, McGraw-Hill Book Company.

Groves, M. O. (2005), 'How Important is Your Personality? Labor Market Returns to Personality for Women in the US and UK', *Journal of Economic Psychology* **26**: 827-841.

Hair Jr, J. F., R. E. Anderson, Tatham, R. L. and Black, W. C. (1998), *Multivariate Data Analysis*, Prentice-Hall International, Inc.

Heckman, J.J. (1978), 'Dummy Endogenous Variables in an Simultaneous Equation System', *Econometrica*, vol. 49, pp. 931-959.

Heckman, J.J. (1979), 'Sample Selection Bias as a Specification Error', *Econometrica*, vol. 47, pp. 153-161.

Hsiao, C. (2003), *Analysis of Panel Data* Cambridge, Cambridge University Press.

Isacsson, G. (2007). 'Twin Data vs. Longitudinal Data to Control for Unobserved Variables in Earnings Functions - Which Are the Differences?', *Oxford Bulletin of Economics and Statistics* **69**(3): 339-362.

Kennedy, P. (1998), *A Guide to Econometrics*. Oxford, Blackwell Publishers.

Leamer, E. E. (1978), *Specification Searches: Ad Hoc Inference with Non-experimental Data*. New York, John Wiley.

Lester, L. H. (2007), 'Immigrant Labour Market Success: An Analysis of the Index of Labour Market Success', *Working Paper 159*, National Institute of Labour Studies, Adelaide.

Lester, L. H. (2008), 'Measuring, Modelling, and Monitoring the Dynamics of Labour Market Success and Successful Settlement of Immigrants to Australia', *National Institute of Labour Studies*, Flinders University, Adelaide, PhD Thesis.

Ligon, E. (2002), 'Dynamic Bargaining in Households', Mimeo, UC Berkeley.

Maas, C. J. M. and J. J. Hox (2004), 'Robustness Issues in Multilevel Regression Analysis', *Statistica Neerlandica* **58**(2): 127-137.

Mazerolle, M. J. and G. Singh (2004), 'Economic and Social Correlates of Re-Employment Following Job Displacement', *The American Journal of Economics and Sociology* **63**(3): 717-730.

Mitchell, W. and A. Bill (2005), 'A Spatial Econometric Analysis of the Irreversibility of Long Term Unemployment in Australia', *Working Paper 05-05*. Newcastle, Centre of Full Employment and Equity.

Montenegro, A. (2001), 'On Sample Size and Precision in Ordinary Least Squares', *Journal of Applied Statistics* **28**(5): 603-605.

Nijman, T. and Verbeek, M. (1992), 'Non-response in Panel Data: The impact on estimates of a life-cycle consumption function', *Journal of Applied Econometrics*, vol. 7, pp. 243-257.

Pissarides, C. (1976), *Labour Market Adjustment*. Cambridge, Cambridge University Press.

Productivity Commission. (2005). Economic Implications of an Ageing Australia, Technical Papers. Canberra.

Richardson, S., (1998), 'Who Gets Minimum Wages?', *Journal of Industrial Relations*, 40(4): 554-579.

Richardson, S. and Harding, A., (1999), 'Poor Workers?', in S Richardson (ed), *Reshaping the Labour Market*, Cambridge University Press, Melbourne.

Richardson, S. and R. Teese (2006), ‘A Well-Skilled Future’, *A Well-Skilled Future Tailoring VET to the Emerging Labour Market*. Melbourne, NCVER.

Ridder, G. (1990), ‘Attrition in multi-wave panel data’, in Hartog, J., Ridder, G., Theeuwes, J. (Eds.), *Panel Data and Labour Market Studies*, Elsevier, Amsterdam.

Skrondal, A. and S. Rabe-Hasketh (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, Chapman andHall/CRC.

Tabachnick, B. G. and L. S. Fidell (1996), *Using Multivariate Statistics*. New York, Harper Collins.

Tan, Y., Lester, L. H. and Richardson, S. (2008)’ ‘Labour Force Projections: A Case Study of the Greater Metropolitan Area of New South Wales’, *Australian Bulletin of Labour* **34**(1): 79-99.

Tan, Y. and S. Richardson (2006), ‘Demographic Impacts on the Future Supply of Vocational Skills’, *A Well-Skilled Future Tailoring VET to the Emerging Labour Market*. Melbourne, NCVER.

Todd, T and Eveline, J. (2004), ‘Paper on the Review of the Gender Pay Gap in Western Australia’.

van Klaveren, C. (2008). ‘A Public Good Version of the Collective Household Model’, *Discussion Paper* TI 2008-018/3. Amsterdam, Tinbergen Institute.

Vella, F. and Verbeek, M. (1999), ‘Two-step Estimation of Panel Data Models with Censored Endogenous Variables and Selection Bias’, *Journal of Econometrics*, vol. 90, pp. 239-263.

Vermeulen, F. (2005), ‘And the winner is... An empirical evaluation of unitary and collective labour supply models’, *Empirical Economics* vol. 30: 711–734

Vermeulen, F. (2006), ‘A collective model for female labour supply with non-participation and taxation’, Journal of Population Economics vol. 19: 99–118.

Wachter, M. (1974), ‘A New Approach to the Equilibrium Labour Force’, *Economica, New Series* **41**(161): 35-51.

Winkelmann, R. (2006), ‘Unemployment, Social Capital, and Subjective Well-Being. Bonn’, IZA Discussion Paper No. 2346.

Wooden, M. (1993), *Underemployment, Hidden Unemployment and Immigrants*. Canberra, Australian Government Publishing Services.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, The MIT Press.

Appendix I—Limited Dependent Variable (Probit) Employment and Participation Equations

An individual's probability of being a labour market participant or of being employed is a function of their attributes (and control variables such as current labour market conditions). The probability of the *i*-th individual being a participant or being employed (P = 1) can be written as the nonlinear (logit or probit) function (see e.g. Winkelmann and Winkelmann 1998):

$$\text{Prob}[P_t = 1 \mid X = x] = \Phi[X'\beta] \tag{1}$$

where Φ is the cumulative distribution function of the standard normal distribution.

The estimated coefficients take the form:

$$\frac{\partial \Pr[y_i = 1 \mid X_i]}{\partial x_i} = \Phi(\beta' x_i)\beta \tag{2}$$

The probit model can be represented as the linear model:

$$\text{Prob(Participation)} = \beta_1 x_{1it} + \beta_2 x_{2it} + \cdots + \beta_k x_{ikt} + \varepsilon_{it} \tag{3}$$

In this representation, the left-hand-side of the specification is the probability of being a labour force participant or of being employed to the probably of not being a participant or employed) of being a participant or being employed—functions of the individual's attributes (**X**) (and a random error term).

NILS

National Institute of Labour Studies
Flinders University, Adelaide, Australia