

# Modeling Employment Dynamics with State Dependence and Unobserved Heterogeneity

VICTORIA PROWSE \*

Department of Economics and Nuffield College, University of Oxford

This version: May 2007

## Abstract

This paper considers the problem of determining the extent of any state dependencies in women's labor supply behavior. Employment outcomes are modeled using a dynamic multinomial choice framework including persistent unobserved heterogeneity with a relatively general distribution. In order to ensure reliable parameter estimates, appropriate restrictions are imposed on the distribution of unobservables. Significant state dependence is present in both full-time and part-time employment. State dependencies are overestimated if persistent unobservables are ignored, and underestimated if an overly restrictive form of persistence is imposed.

**KEY WORDS:** Discrete Labor Supply, Unobserved Heterogeneity, Repeated Multinomial Choice.

**JEL CLASSIFICATION:** C15; C25; J6; J22.

---

\*Address for correspondence: Department of Economics, Manor Road Building, Manor Road, Oxford, OX1 3UQ, United Kingdom. Email: victoria.prowse@economics.ox.ac.uk. Telephone: +44 7761447346. I express my thanks for useful comments and discussion of this paper and earlier drafts to the following people: Steve Bond; Kathryn Graddy; Andrey Launov; Valérie Lechene; and Bent Nielsen. I have also benefited from discussions with participants at seminars in Oxford, Surrey and Sussex, at the 2006 European Summer Meeting of the Econometric Society in Vienna and the 2006 Meeting European Meeting of the Econometric Society in Turin. Any remaining errors are my own. Some of the material presented in this paper is developed from Prowse (2005).

# 1 Introduction

There is a great deal of persistence in individuals' labor supply behavior. Persistence is observed both in participation decisions and in the hours of work of those in employment.<sup>1</sup> It is well established that persistence in labor supply behavior can be generated by two different mechanisms. Firstly, individual characteristics may lead an individual to choose repeatedly the same labor market state. Relevant characteristics consist of observables, such as educational attainment and household structure, and individual specific unobservables including unobserved preferences and ability. Alternatively, persistence in labor supply behavior may arise from the effect of an individual's previous labor supply behavior on his or her current labor supply decision. Specifically, as explained by Heckman (1981), an individual's previous labor supply behavior might change prices, preferences or constraints relevant to his or her current labor supply decision. Such effects are often referred to as causal effects and, in a discrete choice setting, the presence of such effects is known as state dependence.

For the purpose of policy evaluation, it is critical to determine the relative contributions of state dependence and individual characteristics to the observed persistence in labor supply behavior. Indeed, if labor supply choices are driven entirely by observed or unobserved individual characteristics then the effect of a policy intervention, such as a wage subsidy or an in-work benefit, will cease the moment the policy is withdrawn. In contrast, if past labor market outcomes exert a causal effect on current labor supply behavior then the policy intervention will have a lasting effect. The dynamic effects of policy interventions clearly have the potential to change vastly the total effect, and consequently desirability, of the policy. The magnitudes of any causal effects of past labor market outcomes on current labor supply behavior are therefore a critical input to policy analysis.

This paper considers the problem of determining the extent of any causal linkages between past and current labor supply behavior. The modeling framework is a dynamic generalization of the discrete choice approach to labor supply that has become popular recently (see, for example, Bingley and Walker, 1997; Duncan and Giles, 1996; Keane and Moffitt, 1998; van Soest, 1995; van Soest *et al.*, 2002). By taking this approach, it is possible to examine the structure of intertemporal dependencies in participation decisions and in the hours of

---

<sup>1</sup>Based on a sample of American women, Francesconi (2002) reports that 87.6% of women who were non-participants last year are also non-participants this year. The corresponding figures for full-time and part-time work are 87.6% and 68.9% respectively. Booth *et al.* (1999) and O'Reilly and Bothfeld (2002) report similar figures for the United Kingdom.

work of those in employment, thus extending the work of Booth *et al.* (1999), Heckman and Borjas (1980), Heckman and Willis (1977) and Hyslop (1999) who focus on the dynamics of labor force participation.

The empirical analysis, which serves both to illustrate the behavior of various estimators and as a study of employment dynamics, is conducted using a longitudinal sample of British women. Bearing in mind the applied setting, three labor market states are differentiated: full-time employment; part-time employment; and non-employment. The quantities of interests are therefore the nature of any causal effects of previously working full-time or part-time or previously being non-employed on an individual's current choice between full-time work, part-time work and non-employment. While a finer definition of the available labor market states could be adopted, this three way categorization is sufficient to capture the primary modes of labor market engagement of British women (see Brendan *et al.*, 1997; Manning and Petrongolo, 2005; O'Reilly and Bothfeld, 2002). Moreover, all of the analysis extends straightforwardly to models with more than three labor market states.

In order to determine whether causal effects are in operation it is necessary to control for unobserved individual characteristics, also referred to as unobserved heterogeneity or simply unobservables, in a sufficiently general fashion. This problem has been explored in detail for the binary model of labor force participation (see, *inter alia*, Heckman and Borjas, 1980; Heckman and Willis, 1977; Hyslop, 1999). The main contribution of this paper is to explore methods of modeling unobserved heterogeneity in a multinomial as opposed to a binary setting. The primary econometric analysis consists of dynamic mixed multinomial logit models, estimated by Maximum Simulated Likelihood. Using this modeling framework, the implications of different assumptions regarding the distribution of unobservables for estimates of quantities of interest, most importantly state dependencies, are explored. The results of dynamic linear probability models, estimated by Generalized Method of Moments, are also presented, and provide a robustness check of the random effects assumption embodied in the dynamic mixed multinomial logit models.

The application of methods for modeling dynamic linear processes to the current dynamic multinomial labor supply problem is symmetric to the use of such methods to study the dynamics of labor force participation (see Hyslop, 1999, for a detailed discussion) or, equivalently, other dynamic binary processes. However, substantial additional complications arise when generalizing a dynamic binary logit (or probit) model to a multinomial setting. Consider a three alternative cross-sectional (or pooled panel data) multinomial

choice problem. There are two errors. The first error represents unobservables that increase the individual's payoff from the first alternative relative to the third alternative, and the second error represents unobservables that increase the individual's payoff from the second alternative relative to the third alternative. In general, it is desirable to permit the two errors to have different variances and to be correlated. Moreover, the variances of the two errors and their covariance are typically unknown and therefore must be estimated.<sup>2</sup> However, Keane (1992) showed that, while formally identified, multinomial choice models in which the distribution of unobservables contains unknown parameters exhibit symptoms of "fragile identification" if there are no explanatory variables that vary across the choice alternatives. Although the current problem is dynamic rather than static, it fits into this category as the available explanatory variables consist solely of individual characteristics and the individual's employment history. The fragile identification problem arises because it is possible to adjust the intercepts and coefficients on explanatory variables so as to offset almost completely the effect on choice probabilities of changes in the parameters of the distribution of unobservables. Parameter estimates, if convergence is obtained, are inconsistent and standard errors can be very large.<sup>3</sup>

This paper argues that the fragile identification problem can be avoided by utilizing the repeated nature of observations of women's employment outcomes together with some relatively mild restrictions on permissible distributions of unobservables. The restrictions require that all parameters appearing in the distribution of unobservables are formally identified solely from the extra information concerning the distribution of unobservables that is obtained by having repeated observations rather than a cross-section of observations. This necessitates a degree of persistence in unobservables. However, many popular specifications of unobservables, including random intercepts, random coefficients and autocorrelated unobservables, are permitted. Monte Carlo simulations demonstrate that dynamic mixed multinomial logit models with distributions of unobservables satisfying this requirement have desirable empirical properties and therefore do not suffer from the fragile identification prob-

---

<sup>2</sup>The joint distribution of the two errors may contain further unknown shape parameters. This possibility is ignored in this discussion for reasons of simplicity. However, the econometric analysis will permit further generality in the distribution of unobservables.

<sup>3</sup>In some multinomial choice problems, alternative specific explanatory variables are available and therefore the fragile identification problem may be avoided. For example, in models of brand choice price and advertising variables differ across products (applications include Keane, 1997, and Chintagunta, 1992), and in studies of travel mode choice price and travel time vary across the different modes (see Ben-Akiva *et al.*, 1993; Bhat, 1998; Brownstone and Train, 1998). Similarly, in studies of voting behavior data on individuals' opinions of each of the parties or candidates are sometimes available (see Lacy and Burden, 1999; Quinn and Martin, 1999).

lem.

Using the above described methodology, dynamic mixed multinomial logit models with various relatively flexible distributions of unobservables are estimated. Irrespective of the specification of unobservables, significant positive own state dependencies are present in both full-time and part-time employment. Thus, temporary policies incentivizing women to move into full-time or part-time work will affect women's employment behavior beyond the duration of the policy. A comparison of the results across the different specifications of unobservables reveals that estimates of state dependencies are sensitive to the assumed distribution of unobservables. As has been frequently found in the literature on labor force participation, state dependencies are overestimated if persistent unobservables are ignored. Less predictably, estimated state dependencies tend to increase as the distribution of unobservables is generalized from a specification allowing time invariant individual specific effects to more general specifications allowing autocorrelated unobservables and/or random coefficients. Dynamic linear probability models allowing time invariant individual specific fixed effects imply state dependencies comparable to those resulting from a dynamic mixed multinomial logit model with time invariant individual specific random effects. This suggests that the random effects assumption embodied in the dynamic mixed multinomial logit models is valid.

The next section outlines a model describing a woman's choice between full-time employment, part-time employment and non-employment in each year of her career. Section 3 discusses two econometric approaches to modeling the dynamics of women's labor supply choices: dynamic mixed multinomial logit models; and dynamic linear probability models. Restrictions on the nature of persistent unobservables permitted in each of the two approaches are outlined, as are assumptions concerning the initial conditions. Drawing on the related literature, the likely performance of the estimators of the parameters of the two models is outlined. Particular attention is paid to the fragile identification problem affecting dynamic mixed multinomial logit models. This discussion concludes with a description of the strategy used to avoid this problem. Section 4 discusses the British Household Panel Survey, the data source used in this application. Section 5 contains the results of the dynamic mixed multinomial logit models and Section 6 presents the results of the dynamic linear probability models. Section 7 concludes the paper with a discussion of the implications of the results and possible applications of the methodology used in this paper to other multinomial choice problems. Appendices I and II are devoted to Monte Carlo simula-

tions illustrating the performance of estimators of the parameters of several dynamic mixed multinomial logit models, while an analysis of the goodness of fit of the dynamic mixed multinomial logit models is presented in Appendix III.

## 2 Model

The model describes a woman's choice between full-time employment ( $f$ ), part-time employment ( $p$ ) and non-employment ( $n$ ) in each year of her career. For reasons explained below, all of the econometric analysis is reduced form rather than structural; the parameters of this model are not directly estimated. However, this model forms the basis of both econometric approaches used in this study, and provides a framework for discussing the various possible sources of persistence in employment outcomes.

The model is as follows. Every year from  $t = 1, \dots, T$  the woman chooses hours of work,  $h_t$ , to maximize the discounted expected value of her utility subject to a year by year budget constraint. Feasible hours of work are restricted to zero, corresponding to non-employment, full-time hours,  $H_f$ , and part-time hours,  $H_p$ .<sup>4</sup> The discounted expected value of the woman's utility in year  $t$  is as follows

$$U_t = \sum_{s=t}^T \delta^{s-t} E_s u(C_s, h_s; \Omega_{s-1}, Z_s), \quad (1)$$

$$= \sum_{s=t}^T \delta^{s-t} E_s \nu(h_s, \Omega_{s-1}, Z_s), \quad (2)$$

where  $\delta$  is the discount factor. In Equation (1)  $u(C_t, h_t; \Omega_{t-1}, Z_t)$  is the woman's per year utility defined over consumption,  $C_t$ , and hours of work. Relevant elements of the woman's employment history,  $\Omega_{t-1}$ , and observed and unobserved individual characteristics,  $Z_t$ , act as preference shifters. Equation (2) represents the woman's choice problem after consumption has been eliminated using the budget constraint. Here,  $\nu(h_t, \Omega_{t-1}, Z_t)$  is a reduced form utility function defined over hours of work, the woman's employment history and individual characteristics. In the reduced form representation, the woman's employment history and

---

<sup>4</sup>Discrete choice labor supply models have three advantages over continuous or discrete-continuous models of labor supply. Firstly, a discrete opportunity set reflects the fact that many individual's face a choice between a small number of wage-hours contracts, and consequently are unable to vary their hours of work continuously. Secondly, the grouping of reported hours into a small number of categories tends to reduce measurement error. Lastly, and most importantly, discrete approaches generate empirically tractable labor supply functions consistent with non-linearities or discontinuities in budget set generated by fixed costs of employment or the structure of the tax and benefit system (see Duncan and Giles, 1996; van Soest, 1995).

individual characteristics enter the model due to their effect on preferences and also their effect on consumption via the budget constraint.

The value function  $V_t(\Omega_{t-1}, Z_t)$  is the maximized discounted expected value of the woman's utility in year  $t$  given relevant elements of her employment history and her individual characteristics.  $V_t(\Omega_{t-1}, Z_t)$  takes the following form

$$V_t(\Omega_{t-1}, Z_t) = \max \begin{bmatrix} \nu(H_f, \Omega_{t-1}, Z_t) + \delta E_t V_{t+1}(H_f, \Omega_{t-1}, Z_t) \\ \nu(H_p, \Omega_{t-1}, Z_t) + \delta E_t V_{t+1}(H_p, \Omega_{t-1}, Z_t) \\ \nu(0, \Omega_{t-1}, Z_t) + \delta E_t V_{t+1}(0, \Omega_{t-1}, Z_t) \end{bmatrix} \quad (3)$$

In the above  $V_{t+1}(h_t, \Omega_{t-1}, Z_t)$  is the woman's value function in year  $t + 1$  given the relevant elements of her employment history in year  $t - 1$  and her decision to work  $h_t$  hours in year  $t$ . A utility maximizing woman will choose full-time employment in year  $t$  if the sum of her contemporaneous payoff from full-time employment and the maximized discounted expected value of her utility in future years in she works full-time in the current year is greater than her corresponding payoffs from both part-time employment and non-employment. The decision criteria for part-time work and non-employment are defined symmetrically.

This formulation is sufficiently general so as to allow dependencies between a woman's past and current labor supply decisions due to habit formation in labor supply behavior (see Bover, 1991; Kubin and Prinz, 2002; Muellbauer, 1988; Woittiez and Kapteyn, 1998), wage based rewards for human capital accumulated via labor market experience (see Altug and Miller, 1998; Eckstein and Wolpin, 1989; Imai and Keane, 2004; Keane and Wolpin, 2001; Shaw, 1989; Wolpin, 1992) and job search costs (see Heckman and Borjas, 1980; Hyslop, 1999; Layard and Bean, 1989). Habit formation gives rise to non-separabilities in the utility function, while accumulation of human capital and job search costs imply non-separabilities in the budget constraint. All three mechanisms create dependencies between past and current labor supply choices. More specifically, job search costs generate dependencies between labor supply choices in consecutive years, while habit formation and accumulation of human capital have the potential to create dependencies in labor supply behavior spanning several years.

### 3 Empirical Implementation

A woman's employment choice depends on her individual characteristics,  $Z_t$ . However, only a subset of these characteristics, denoted  $X_t$ , are observed by the econometrician. Conditional on observables, the probabilities of the woman choosing each of the three employment states are as follows

$$P_{f,t}(\Omega_{t-1}, X_t) = P \left( \begin{array}{l} V_t^f(\Omega_{t-1}, Z_t) \geq V_t^p(\Omega_{t-1}, Z_t) \\ V_t^f(\Omega_{t-1}, Z_t) \geq V_t^n(\Omega_{t-1}, Z_t) \end{array} \right), \quad (4)$$

$$P_{p,t}(\Omega_{t-1}, X_t) = P \left( \begin{array}{l} V_t^p(\Omega_{t-1}, Z_t) > V_t^f(\Omega_{t-1}, Z_t) \\ V_t^p(\Omega_{t-1}, Z_t) \geq V_t^n(\Omega_{t-1}, Z_t) \end{array} \right), \quad (5)$$

$$P_{n,t}(\Omega_{t-1}, X_t) = P \left( \begin{array}{l} V_t^n(\Omega_{t-1}, Z_t) > V_t^f(\Omega_{t-1}, Z_t) \\ V_t^n(\Omega_{t-1}, Z_t) > V_t^p(\Omega_{t-1}, Z_t) \end{array} \right), \quad (6)$$

where  $P()$  denotes a probability,  $V^j(\Omega_{t-1}, Z_t) = \nu(H_j, \Omega_{t-1}, Z_t) + \delta E_t V_{t+1}(H_j, \Omega_{t-1}, Z_t)$  for  $j = f, p$  and  $V^n(\Omega_{t-1}, Z_t) = \nu(0, \Omega_{t-1}, Z_t) + \delta E_t V_{t+1}(0, \Omega_{t-1}, Z_t)$  (see Equation (3)).

Estimation of the parameters of the above dynamic programming problem would be feasible in the absence of persistent unobserved heterogeneity. However, computational difficulties are encountered when estimating this type of model if controls for persistent unobserved individual characteristics are included (see Eckstein and Wolpin, 1989; Keane and Wolpin, 1997; Rust, 1996). Given this study aims to separate state dependence from the effects of persistent unobserved individual characteristics, it seems that working directly with the above dynamic programming problem is not desirable.

In contrast, by making appropriate approximations it is possible to explore the contribution of unobserved individual characteristics to the observed persistence in women's labor supply behavior. The main econometric analysis proceeds by making approximations to the value functions. Placing appropriate distributional assumptions on unobservables gives rise to dynamic mixed multinomial logit models. In order to check the robustness of the results to assumptions placed on the distribution of unobservables, dynamic linear probability models are also estimated. These models may be viewed as approximations to choice probabilities rather than value functions. The two econometric approaches, together with their relative merits, are described below.



### 3.1 Dynamic Mixed Multinomial Logit Models

An examination of Equations (4)-(6) reveals that labor supply probabilities can be expressed in terms of the two indices  $V_t^f(\Omega_{t-1}, Z_t) - V_t^n(\Omega_{t-1}, Z_t)$  and  $V_t^p(\Omega_{t-1}, Z_t) - V_t^n(\Omega_{t-1}, Z_t)$ .<sup>5</sup> The former difference in value functions describes the woman's payoff from full-time work relative to non-employment while the latter difference in value functions describes the woman's payoff from part-time work relative to non-employment. The following approximations are adopted

$$V_t^f(\Omega_{t-1}, Z_t) - V_t^n(\Omega_{t-1}, Z_t) = X_{i,t}b_f + \gamma_f\Omega_{i,t-1} + \varrho_{i,f,t}, \quad (7)$$

$$V_t^p(\Omega_{t-1}, Z_t) - V_t^n(\Omega_{t-1}, Z_t) = X_{i,t}b_p + \gamma_p\Omega_{i,t-1} + \varrho_{i,p,t}. \quad (8)$$

The subscript  $i$  denotes the individual under consideration. The coefficient vectors  $\gamma_f$  and  $\gamma_p$  measure the effect of the woman's employment history on her payoff from full-time employment relative to her payoff from non-employment and on her payoff from part-time employment relative to her payoff from non-employment. State dependencies are present if any elements of  $\gamma_f$  or  $\gamma_p$  are significantly different from zero. The econometric analysis is conducted using panel data where information about a woman's employment history is restricted to the duration of the woman's presence in the panel. Thus, prior to estimation, restrictions on the specification of  $\Omega_{i,t-1}$  are required. Attention is restricted to the case where only the woman's labor market outcomes in the past two years affect her current period payoffs. Specifically  $\Omega_{i,t-1} = [Y_{i,f,t-1}, Y_{i,p,t-1}, Y_{i,f,t-2}, Y_{i,p,t-2}]$ , where  $Y_{i,j,t}$  is an indicator taking the value one if woman  $i$  was in employment state  $j$  at time  $t$  and zero otherwise. Equations (7) and (8) then hold for  $t = 3, \dots, T_i$  where  $t = 1$  and  $t = T_i$  are, respectively, the first and last observations of woman  $i$ 's employment outcomes. This specification should not be overly restrictive as the strongest intertemporal dependencies in labor supply incentives are likely to occur over short time horizons. Further support for this specification arises from the results for the dynamic linear probability models; in a linear setting earlier lags of employment outcomes are found to be insignificant.<sup>6</sup>

The coefficient vectors  $b_f$  and  $b_p$  measure the effects of the individual characteristics in  $X_{i,t}$ , such as educational qualifications and household structure variables, on a woman's payoffs from full-time and part-time work relative to her payoff from non-employment.  $X_{i,t}$  does not include employment state specific wages or incomes because such quantities are

<sup>5</sup>The third index  $V_t^f(\Omega_{t-1}, Z_t) - V_t^p(\Omega_{t-1}, Z_t)$  is redundant as it is equal to the difference between the second and first indices.

<sup>6</sup>See Table 10.

unobserved for all states not chosen by the woman. Coefficients therefore measure both differences in labor supply preferences according to observed individual characteristics and the effect of observed individual characteristics on alternative specific wages and incomes. It is further noted that, in general, a woman's characteristics,  $X_{i,t}$ , and her employment history,  $\Omega_{i,t-1}$ , have both a direct effect on her current period employment choice, via the woman's current period payoff, and an indirect effect, via the woman's payoffs from future employment decisions. Coefficient estimates should therefore be interpreted as the combined effects of individual characteristics on a woman's current and future payoffs.

$\varrho_{i,f,t}$  and  $\varrho_{i,p,t}$  denote all unobservables relevant to the woman's employment decision in year  $t$ . Dynamic mixed multinomial logit models are obtained by placing appropriate distributional assumptions on unobservables. The distributional assumptions restrict unobservables to be random effects, occurring independently of  $X_{i,t}$ . However, within the random effects framework, relatively general intra- and intertemporal distributions of unobservables are permitted.

For the purpose of deriving choice probabilities  $\varrho_{i,f,t}$  and  $\varrho_{i,p,t}$  are decomposed into two components. Specifically, for  $t = 3, \dots, T_i$

$$\varrho_{i,f,t} = \eta_{i,f,t} + \xi_{i,f,t}, \quad (9)$$

$$\varrho_{i,p,t} = \eta_{i,p,t} + \xi_{i,p,t}. \quad (10)$$

Generality in the distribution of unobservables enters the model through the joint distribution of  $\eta_{i,f,t}$  and  $\eta_{i,p,t}$ . Restrictive distributional assumptions are placed on  $\xi_{i,f,t}$  and  $\xi_{i,p,t}$ . In particular

- A1**  $\xi_{i,f,t}$  and  $\xi_{i,p,t}$  are defined respectively as  $\epsilon_{i,1,t} - \epsilon_{i,3,t}$  and  $\epsilon_{i,2,t} - \epsilon_{i,3,t}$  for  $t = 3, \dots, T_i$ .
- A2**  $\epsilon_{i,j,t}$  for  $j = 1, 2, 3$  and  $t = 3, \dots, T_i$  are mutually independent.
- A3**  $\epsilon_{i,j,t}$  are independent of  $X_{i,s}$  for  $j = 1, 2, 3$  and  $s, t = 3, \dots, T_i$ .
- A4**  $\epsilon_{i,j,t}$  for  $j = 1, 2, 3$  and  $t = 3, \dots, T_i$  have type I extreme value distributions implying

$$P(\epsilon_{i,j,t} \leq q) = \exp(-\exp(-q)).$$

Let  $\eta_{i,f}$  and  $\eta_{i,p}$  denote, respectively,  $\eta_{i,f,t}$  and  $\eta_{i,p,t}$  stacked over  $t$ . Assumptions **A1-A4** imply that conditional on  $(\eta_{i,f}, \eta_{i,p}, X_{i,t}, \Omega_{i,t-1})$  the woman's choice probabilities at  $t = 3, \dots, T_i$

are independent over time and take the familiar multinomial logit form

$$P_{i,f,t}(X_{i,t}, \Omega_{i,t-1} | \eta_{i,f}, \eta_{i,p}) = \frac{\exp(X_{i,t}b_f + \gamma_f\Omega_{i,t-1} + \eta_{i,f,t})}{1 + \sum_{j=f,p} \exp(X_{i,t}b_j + \gamma_j\Omega_{i,t-1} + \eta_{i,j,t})}, \quad (11)$$

$$P_{i,p,t}(X_{i,t}, \Omega_{i,t-1} | \eta_{i,f}, \eta_{i,p}) = \frac{\exp(X_{i,t}b_p + \gamma_p\Omega_{i,t-1} + \eta_{i,p,t})}{1 + \sum_{j=f,p} \exp(X_{i,t}b_j + \gamma_j\Omega_{i,t-1} + \eta_{i,j,t})}, \quad (12)$$

$$P_{i,n,t}(X_{i,t}, \Omega_{i,t-1} | \eta_{i,f}, \eta_{i,p}) = \frac{1}{1 + \sum_{j=f,p} \exp(X_{i,t}b_j + \gamma_j\Omega_{i,t-1} + \eta_{i,j,t})}. \quad (13)$$

### 3.1.1 Treatment of the Initial Conditions

An initial conditions problem arises when estimating this model. Given the dynamic structure of the model, a woman's employment outcome in the year  $t = 1$  depends on her unobserved employment outcomes in the years  $t = 0$  and  $t = -1$ . Likewise, the woman's employment outcome in the year  $t = 2$  depends on her unobserved employment outcome in the year  $t = 0$ . Therefore, employment outcomes in the years  $t = 1$  and  $t = 2$ , referred to as the initial conditions, cannot be modeled in the same way as subsequent employment outcomes.

The treatment of the initial conditions proposed by Wooldridge (2005) is adopted. This approach involves forming a likelihood contribution for each woman that is conditional on the initial conditions. In more detail, individual likelihood contributions are formed by integrating the joint probability of a woman's observed employment outcomes for  $t = 3, \dots, T_i$  conditional on unobserved heterogeneity, exogenous explanatory variables and the woman's first two observed employment outcomes with respect to the distribution of unobserved heterogeneity conditional on exogenous explanatory variables and the woman's employment behavior in her first two years in the sample.<sup>7</sup> Let  $G(\eta_{i,f}, \eta_{i,p} | X_i, IC_i)$  denote the distribution of  $(\eta_{i,f}, \eta_{i,p})$  given the collection of all exogenous explanatory variables,  $X_i$ , and the woman's

---

<sup>7</sup>In contrast, Heckman (1981) suggests modeling the joint distribution of a woman's employment outcomes over the entire sample period. This requires a specification of the joint distribution of the woman's employment outcomes in the years  $t = 1$  and  $t = 2$  conditional on exogenous explanatory variables and unobserved heterogeneity, and a specification of the distribution of unobserved heterogeneity. In general, the exact distribution of the initial conditional is impossible to derive. (In the absence of non-stationary explanatory variables it is possible to derive the equilibrium distribution of the process which can then be used as the distribution of the initial conditions. However the presence of a number of non-stationary explanatory variables, including age and the time dummies, in the current application renders using an exact specification of the initial conditions impossible.) Instead, Heckman (1981) suggests approximating the distribution of the initial observations conditional on unobserved heterogeneity and exogenous explanatory variables. The Wooldridge approach does not require a model of the joint distribution of the woman's employment outcomes in the years  $t = 1$  and  $t = 2$  or a specification of the joint distribution of unobservables occurring at  $t = 1$  and  $t = 2$  and the unobservables occurring in future years. This has the computational advantage, relative to the Heckman (1981) approach, of reducing the number of unknown parameters.

employment outcomes in the years  $t = 1$  and  $t = 2$ ,  $IC_i$ .<sup>8</sup> Woman  $i$ 's likelihood contribution is as follows

$$P_i = \int \prod_{t=3}^{T_i} \prod_{j=f,p,n} P_{i,j,t}(X_{i,t}, \Omega_{i,t-1} | \eta_{i,f}, \eta_{i,p})^{Y_{i,j,t}} dG(\eta_{i,f}, \eta_{i,p} | X_i, IC_i) \quad (14)$$

The likelihood function for a sample of  $N$  women whose labor market outcomes are assumed to be independent is given by

$$\mathcal{L} = \prod_{i=1}^N \int \prod_{t=3}^{T_i} \prod_{j=f,p,n} P_{i,j,t}(X_{i,t}, \Omega_{i,t-1} | \eta_{i,f}, \eta_{i,p})^{Y_{i,j,t}} dG(\eta_{i,f}, \eta_{i,p} | X_i, IC_i). \quad (15)$$

Proceeding further requires assumptions regarding the distribution function  $G(\eta_{i,f}, \eta_{i,p} | X_i, IC_i)$ .

### 3.1.2 The Distribution of Unobservables

The distribution function  $G(\eta_{i,f}, \eta_{i,p} | X_i, IC_i)$  dictates the structure of persistence in unobservables and the joint distribution of unobservables occurring in a particular year. As discussed above, allowing persistence in unobservables is necessary to determine correctly the nature of state dependencies in labor supply behavior. Furthermore, Chintagunta (1992) and Hausman and Wise (1979) show that estimates of quantities of interest including marginal effects, substitution patterns and elasticities are not robust to the intratemporal distribution of unobservables. It is therefore desirable to work with a distribution of unobservables that is not overly restrictive.

The model is estimated with several different specifications of unobservables, described below in Section 5. All of the specifications are nested within the following structure

$$\eta_{i,f,t} = \omega_{i,f} \Omega_{i,t-1} + \pi_{i,f} W_{i,t} + (\vartheta_f + \psi_{i,f}) IC_i + \varpi_{i,f,t} \quad \text{for } t = 3, \dots, T_i, \quad (16)$$

$$\eta_{i,p,t} = \omega_{i,p} \Omega_{i,t-1} + \pi_{i,p} W_{i,t} + (\vartheta_p + \psi_{i,p}) IC_i + \varpi_{i,p,t} \quad \text{for } t = 3, \dots, T_i, \quad (17)$$

where  $\varpi_{i,f,t}$  and  $\varpi_{i,p,t}$  contain both time invariant components and first order autoregressive

---

<sup>8</sup>In the empirical implementation,  $IC_i$  is a vector of indicator variables detailing whether the woman was working full-time in both  $t = 1$  and  $t = 2$ , working part-time in both  $t = 1$  and  $t = 2$ , was non-employed in both  $t = 1$  and  $t = 2$ , worked both full-time and part-time in her first two years in the sample, worked full-time and was non-employed in her first two years in the sample or worked part-time and was non-employed in her first two years in the sample.

processes

$$\overline{\omega}_{i,f,t} = \nu_{i,f} + \epsilon_{i,f,t} \quad \text{where} \quad \epsilon_{i,f,t} = \rho_f \epsilon_{i,f,t-1} + \xi_{i,f,t}, \quad (18)$$

$$\overline{\omega}_{i,p,t} = \nu_{i,p} + \epsilon_{i,p,t} \quad \text{where} \quad \epsilon_{i,p,t} = \rho_p \epsilon_{i,p,t-1} + \xi_{i,p,t}. \quad (19)$$

$\eta_{i,f,t}$  and  $\eta_{i,p,t}$  consist of four distinct components: (i)  $\omega_{i,f}$  and  $\omega_{i,p}$  are the random components of the coefficients on the woman's employment history; (ii)  $\pi_{i,f}$  and  $\pi_{i,p}$  represent the random components of the coefficients on the explanatory variables,  $W_{i,t}$ ; (iii)  $\psi_{i,f}$  and  $\psi_{i,p}$  are the random components of the coefficients on the initial conditions and  $\vartheta_f$  and  $\vartheta_p$  are the deterministic component of the coefficients on the initial conditions; and (iv)  $\overline{\omega}_{i,f,t}$  and  $\overline{\omega}_{i,p,t}$  represent the random components of the intercepts.

The random intercepts represent unobserved differences in women's payoffs that occur irrespective of past employment outcomes or individual characteristics. Women with different values of the random intercepts have intrinsic differences in the unobserved components of their payoffs. Such differences may result from, for example, systematic differences in preferences or ability differences. Persistence over time in the random intercepts, either through the time invariant or autocorrelated components of the random intercepts, provides one source of persistent unobservables that may contribute to persistence in employment outcomes. Intratemporal heteroscedasticity in the random intercepts implies different amounts of unobserved variation in women's preferences for full-time and part-time employment relative to non-employment. Similarly, intratemporal correlation between the random intercepts allows positive or negative association in women's unobserved preferences for full-time and part-time employment relative to non-employment. Intratemporal heteroscedasticity and correlations in the random intercepts break the cross-sectional *i.i.d* structure of unobservables and thus allow flexibility in within period substitution patterns.

The random coefficients on the explanatory variables allow women to have different unobserved preferences depending on their observed characteristics. In the empirical implementation the random coefficients are used to allow a higher level of variation in a woman's unobserved preferences if she has a young child or a degree level qualification. High levels of variation in unobserved preferences among women with young children reflects variation in child-care costs, while women with a degree level qualification will have high levels of unobserved variation in their payoffs if there is a relatively high level of heterogeneity in the labor market returns to a university education.

The random coefficients on the lagged dependent variables allow different distributions of unobservables depending on employment behavior in the previous two years. Higher variation in unobserved preferences for working full-time among women already working full-time than among women working part-time or in non-employment is permitted. Such a distribution of unobservables will arise if women have different propensities to capitalize on labor market experience, face different costs of job search or are subject to different rates of habit formation. The coefficients on the initial conditions allow differences in unobservables according to a woman's labor market outcomes in her first two years in the sample.

In the empirical implementation, the time invariant components of the random intercepts  $(v_{i,f}, v_{i,p})$  are, depending on the specification of unobservables under consideration, either jointly normally distributed with mean zero and an unrestricted covariance matrix or have a distribution generated by a mixture of two normal distributions. All other pairs of random coefficients, for example the  $k^{\text{th}}$  elements of  $\pi_{i,f}$  and  $\pi_{i,p}$ , are jointly normally distributed with zero mean and an unrestricted covariance matrix, and all pairs of random coefficients are mutually independent.<sup>9</sup> In the specifications allowing autocorrelated intercepts,  $\xi_{i,f,t-1}$  and  $\xi_{i,p,t-1}$  are assumed to be jointly normally distributed with variances such that the autocorrelation processes are stationary.

### 3.1.3 Estimation

Analytic expressions for the likelihood function given above in Equation (15) are unavailable for all but the simplest specifications of unobserved heterogeneity. In problems where the dimension of integration is three or more, simulation techniques are the most appropriate method of evaluating the likelihood contributions.<sup>10</sup> Simulation methods, of which there are many variants, replace the intractable integral in the likelihood function by a sum over likelihood functions evaluated at different draws from the distribution of unobserved heterogeneity.

Let  $(\eta_{i,f}^r, \eta_{i,p}^r)$  denote the  $r^{\text{th}}$  draw from the distribution  $G(\eta_{i,f}, \eta_{i,p} | X_i, IC_i)$  for individual

---

<sup>9</sup>Due to the high levels of persistence in employment outcomes and the persistence in many explanatory variables, it is difficult to estimate the covariances between a large number of random coefficients.

<sup>10</sup>For two dimensional problems fast and accurate cubature methods are available to evaluate the individual likelihood contributions (Geweke, 1996, provides a survey). However numerical methods are unable to evaluate the likelihood contributions with sufficient speed and accuracy to be effective in problems where the dimension of integration is greater than two (see Bhat, 2001; Hajivassiliou and Rudd, 1994).

*i.* The likelihood is simulated as follows

$$\mathcal{L}_s = \frac{1}{R} \prod_{i=1}^N \sum_{r=1}^R \prod_{t=3}^{T_i} \prod_{j=f,p,n} P_{i,j,t}(X_{i,t}, \Omega_{i,t-1} | \eta_{i,f}^r, \eta_{i,p}^r)^{Y_{i,j,t}}. \quad (20)$$

Maximum Simulated Likelihood estimates are obtained by maximizing the log simulated likelihood function. By the strong law of large numbers the Maximum Simulated Likelihood estimates converge almost surely to the true parameters as  $R \rightarrow \infty$  and  $N \rightarrow \infty$ . Moreover, if  $R$  increases at a fast enough rate relative to  $N$ , Maximum Simulated Likelihood estimation is asymptotically equivalent to Maximum Likelihood estimation. In particular, with pseudo random draws the rate of convergence is  $R^{-0.5}$  (Hajivassiliou and Rudd, 1994).

In this application, Maximum Simulated Likelihood estimates are obtained using antithetic variates rather than pseudo random draws.  $R$  antithetic draws are obtained by taking  $R/2$  independent draws from the distribution of  $(\eta_{i,f}, \eta_{i,p})$ , denoted  $\{\eta_{i,f}^r, \eta_{i,p}^r\}_{r=1}^{R/2}$ . The remaining  $R/2$  draws are given by  $\{2\mu - (\eta_{i,f}^r, \eta_{i,p}^r)\}_{r=1}^{R/2}$ , where  $\mu$  is the mean of  $(\eta_{i,f}, \eta_{i,p})$ . Hajivassiliou (1999) shows that the use of antithetic variates in Maximum Simulated Likelihood problems approximately halves the number of draws required to obtain a given level of accuracy. Monte Carlo simulations presented in Appendix II show that Maximum Simulated Likelihood estimation with 5000 antithetic draws yields estimates with tolerably small amounts of bias, and therefore  $R = 5000$  is used in thought the empirical analysis.

### 3.2 Dynamic Linear Probability Models

The second econometric approach approximates choice probabilities by linear functions of observed individual characteristics and past employment outcomes. This approximation results in dynamic linear models for the probabilities of full-time and part-time work. The linearity of these models means that results will not necessarily be consistent with an interpretation as choice probabilities. Furthermore, the approximation does not recognize the multinomial structure of the model. However, the linearity of the models allows the inclusion of persistent unobserved characteristics with an unrestricted distribution and an unrestricted relationship with the explanatory variables. The estimation results from these models therefore provide a robustness check for the results from the dynamic mixed multinomial logit models where all unobservables are assumed to have specific distributions and to be independent of the explanatory variables.

The dynamic linear probability models take the form of two equations

$$Y_{i,f,t} = X_{i,t}\beta_f + \lambda_f\Omega_{i,t-1} + \theta_{i,f} + \varepsilon_{i,f,t} \text{ for } t = 3, \dots, T_i, \quad (21)$$

$$Y_{i,p,t} = X_{i,t}\beta_p + \lambda_p\Omega_{i,t-1} + \theta_{i,p} + \varepsilon_{i,p,t} \text{ for } t = 3, \dots, T_i. \quad (22)$$

In the above  $\beta_f$ ,  $\beta_p$ ,  $\lambda_f$  and  $\lambda_p$  are suitably dimensioned vectors of unknown parameters. Total unobservables have been decomposed into time invariant components,  $\theta_{i,f}$  and  $\theta_{i,p}$ , and time-varying components  $\varepsilon_{i,f,t}$  and  $\varepsilon_{i,p,t}$ . The specification of  $\Omega_{i,t-1}$  is as described above in Section 3.1. As in the dynamic mixed multinomial logit model presented above, employment state specific wages or incomes are not included. Coefficient estimates should be therefore be interpreted as the effects of individual characteristics on labor supply behavior occurring through both preferences and alternative specific wages and incomes. Also, as above, coefficients reflect the effects of individual characteristics on both contemporaneous and expected future payoffs.

Throughout the analysis two sets of assumptions on  $\varepsilon_{i,j,t}$  are maintained

$$E[\varepsilon_{i,j,t}\varepsilon_{i,j,s}] = 0 \text{ for all } s, t = 3, \dots, T_i \text{ and } j = f, p. \quad (23)$$

$$E[Y_{i,f,1}\varepsilon_{i,j,t}] = E[Y_{i,p,1}\varepsilon_{i,j,t}] = 0 \text{ for } t = 2, \dots, T_i, \text{ and} \quad (24)$$

$$E[Y_{i,f,2}\varepsilon_{i,j,t}] = E[Y_{i,p,2}\varepsilon_{i,j,t}] = 0 \text{ for } t = 3, \dots, T_i. \quad (25)$$

Equation (23) imposes zero serial correlation on  $\varepsilon_{i,j,t}$ . Extensions to allow  $\varepsilon_{i,j,t}$  to have a moving average process or to be autocorrelated are straightforward. However, these generalizations are not presented because specification tests, reported below in Table 10, do not reject the hypothesis that  $\varepsilon_{i,j,t}$  is serially uncorrelated. Equations (24) and (25) are standard moment restrictions on the initial conditions.

Zero serial correlation in  $\varepsilon_{i,j,t}$  together with the binary structure of  $Y_{i,f,t}$  and  $Y_{i,p,t}$  implies that Equations (21) and (22) can be interpreted as linear probability models. In particular, for  $t = 3, \dots, T_i$

$$P(Y_{i,f,t} = 1 | X_{i,t}, \Omega_{i,t-1}, \theta_{i,f}) = E(Y_{i,f,t} | X_{i,t}, \Omega_{i,t-1}, \theta_{i,f}) = X_{i,t}\beta_f + \lambda_f\Omega_{i,t-1} + \theta_{i,f}, \quad (26)$$

$$P(Y_{i,p,t} = 1 | X_{i,t}, \Omega_{i,t-1}, \theta_{i,p}) = E(Y_{i,p,t} | X_{i,t}, \Omega_{i,t-1}, \theta_{i,p}) = X_{i,t}\beta_p + \lambda_p\Omega_{i,t-1} + \theta_{i,p}. \quad (27)$$

Given the linear structure of the model, coefficients can be interpreted as the marginal effects of changes in individual characteristics or past employment outcomes on the woman's current



probabilities of engaging in full-time or part-time work.

When estimating the parameters of Equations (21) and (22) it is possible to allow the time invariant individual effects,  $\theta_{i,f}$  and  $\theta_{i,p}$ , to be correlated with observed individual characteristics,  $X_{i,t}$ , i.e., to be fixed effects rather than random effects.<sup>11</sup> In the presence of individual specific fixed effects consistent estimates of  $[\beta_f, \lambda_f]$  and  $[\beta_p, \lambda_p]$  can be obtained by taking first differences of Equations (21) and (22) and using appropriate lagged levels of the woman's employment history and individual characteristics as instruments for the first difference equations (see Anderson and Hsiao, 1982; Arellano and Bond, 1991; Holtz-Eakin *et al.*, 1988).

Taking first differences of Equations (21) and (22) gives

$$\Delta Y_{i,f,t} = \Delta X_{i,t} \beta_f + \lambda_f \Delta \Omega_{t-1} + \Delta \varepsilon_{i,f,t}, \text{ for } t = 4, \dots, T_i, \quad (28)$$

$$\Delta Y_{i,p,t} = \Delta X_{i,t} \beta_p + \lambda_p \Delta \Omega_{t-1} + \Delta \varepsilon_{i,p,t}, \text{ for } t = 4, \dots, T_i. \quad (29)$$

Given the above assumptions,  $Y_{i,f,t-2}$  and  $Y_{i,p,t-2}$  and earlier lags of these variables are valid instruments for  $\Delta \varepsilon_{i,f,t}$  and  $\Delta \varepsilon_{i,p,t}$ . Under the assumption of strict exogeneity of individual characteristics,  $X_{i,s}$  for all  $s$  are valid instruments for  $\Delta \varepsilon_{i,f,t}$  and  $\Delta \varepsilon_{i,p,t}$ . However, if individual characteristics are predetermined then only individual characteristics dated  $t-1$  and earlier are valid instruments, while if individual characteristics are endogenous only  $X_{i,t-2}$  and earlier values of individual characteristics are valid instruments. Irrespective of whether individual characteristics are strictly exogenous, predetermined or endogenous, the number of moment conditions exceeds the number of parameters provided that  $\max_i \{T_i\} \geq 4$ .<sup>12,13</sup> All available moment conditions can be combined efficiently by using Generalized Method of Moments techniques (see Hansen, 1982). In the presence of over-identifying restrictions, the validity of the instrument set, and of subsets of the instruments, can be tested using a Sargan test.

<sup>11</sup>Due to the dynamic structure of the model,  $\theta_{i,f}$  and  $\theta_{i,p}$  will necessarily be correlated with the woman's employment history,  $\Omega_{i,t-1}$ .

<sup>12</sup>First difference equations are available for  $t = 4, \dots, T_i$ . Therefore, at least four years of data are required to apply this estimator.

<sup>13</sup>Let  $K$  denote the dimension of  $\Delta X_{i,t}$ . It follows that there are  $K+4$  parameters in each of Equations (28) and (29). If  $X_{i,t}$  is endogenous then there are  $(\max_i \{T_i\} - 2)(K+4)$  moment conditions for  $\Delta \varepsilon_{i,f}$  and  $\Delta \varepsilon_{i,p,t}$ . If, instead,  $X_{i,t}$  is predetermined then there are  $K(\max_i \{T_i\} - 1) + 4(\max_i \{T_i\} - 1)$  moment conditions for  $\Delta \varepsilon_{i,f,t}$  and  $\Delta \varepsilon_{i,p}$ .

### 3.3 Identification

The requirements for formal identification of both linear dynamic panel data models and mixed multinomial logit models are well established. For completeness these requirements are outlined below. Previous Monte Carlo simulations and related applications have shown that formal identification provides only a partial guide to the reliability of the estimators of the parameters of these models. This is particularly true in the case of multinomial choice models with non-trivial distributions of unobservables. With this in mind, practical issues surrounding the implementation of these estimators are also discussed.

In linear dynamic panel data models, coefficients on time invariant variables are not identified if the model includes individual specific fixed effects. In light of this and given the very low levels of intertemporal variation in educational qualifications, education related variables are omitted when estimating the first difference Equations (28) and (29). Furthermore, if the underlying autoregressive process, that is either Equation (21) or Equation (22), has a unit root then, given the above describe moment conditions, coefficients on the lagged employment outcomes are not identified. Related to this, the finite sample properties of the estimator have been found to deteriorate as the process approaches a unit root (see Blundell and Bond, 1998). However, given the binary nature of the dependent variables, problems stemming from unit roots are unlikely to be present.

Formal identification of multinomial choice models requires location and scale normalizations (Bolduc, 1992; Bunch and Kitamura, 1991; Bunch, 1991; Dansie, 1985). These requirements arises as individuals' choices are based on differences in payoffs. Location normalizations are required as adding a common increment to each of an individual's payoffs leaves her choice problem unaffected. Similarly, a scale normalization is required as multiplying the payoff associated with each of the alternatives by a common positive constant also leaves the individual's choice problem unaffected. The required location normalizations are already imposed in the above model because the parameterization is in terms of the two indices  $V_t^f(\Omega_{t-1}, Z_t) - V_t^n(\Omega_{t-1}, Z_t)$  and  $V_t^p(\Omega_{t-1}, Z_t) - V_t^n(\Omega_{t-1}, Z_t)$ , rather than the three underlying value functions. In other words, by parameterizing differences in payoffs, rather than the payoffs themselves, the required location normalizations are automatically imposed. Furthermore, in the mixed multinomial logit model described above, a set of scale normalizations are imposed. Specifically, assumption **A4** places a normalization on the variance of the second component of unobservables. Unless otherwise indicated, this is sufficient to ensure scale identification.

Multinomial choice models in which the distribution of unobservables contains unknown parameters often perform poorly in applications where there are no explanatory variables that vary across alternatives, as is the case in the current application where explanatory variables consist of individual characteristics and the individual's previous employment outcomes. The first evidence of this problem was provided by Keane (1992). Using a series of Monte Carlo simulations, Keane (1992) showed that the cross-sectional multinomial probit model with heteroscedastic and correlated unobservables and no alternative specific explanatory variables suffers from a fragile identification problem. The symptoms of fragile identification include possible non-convergence, inconsistent parameter estimates and a close to singular Hessian which translates into huge standard errors. Keane (1992) argues that the fragile identification problem is intrinsic to the structure of multinomial choice models in which the distribution of unobservables contains unknown parameters; in such models it is possible to adjust the intercepts and coefficients on explanatory variables so as to offset almost completely the effect on choice probabilities of changes in the parameters describing the distribution of unobservables. Consequently, the criterion function is almost completely flat over a large subset of the parameter space.

Many researchers cite the results of Keane (1992) as a reason for using a multinomial logit model instead of a model permitting a more general distribution of unobservables (for examples see Adams *et al.*, 1999; Bound *et al.*, 1999; Dancer and Fiebig, 2004; Forman, 2005; Sturm and Wells, 1998). As explained above, imposing the distributional assumptions implicit in the multinomial logit model is likely to produce misleading results. Other researchers, unwilling to accept the restrictions imposed by the multinomial logit model, have sought to construct alternative specific explanatory variables from the available individual specific variables. For example, in a model of schooling choice Giannelli and Monfardini (2003) construct a measure of alternative specific expected future earnings while Heckman and Sedlacek (1985) and Keane (1992) use predicted occupation specific earnings in models of occupational choice. The empirical success of these methods is dependent on sufficient variation across alternatives in the constructed alternative specific variables (see Harris and Keane, 1998). Indeed, Keane (1992) found that in one particular application constructed occupation specific income did not have sufficient variation to solve the fragile identification problem, although the method has been used successfully in other applications.

In contrast to the above described approaches, in the current application the fragile identification problem is avoided by using the extra information contained in repeated ob-

servations of individuals' employment outcomes together with some mild restrictions on the distribution of unobservables. Specifically, attention is restricted to distributions of unobservables in which all parameters describing the intratemporal distribution of unobservables are identified given the joint distribution of unobservables occurring at different times. This requirement ensures that repeated observations yield additional identifying information concerning the parameters of the intratemporal distribution of unobservables. Given that it is the parameters of the intratemporal distribution of unobservables that cause the fragile identification problem, imposing the above described restrictions on the distribution of unobservables is likely to improve the empirical properties of estimators of the parameters of the model.

The required restrictions on the distribution of unobservables are formalized for the case where  $(\eta_{i,f}, \eta_{i,p})$  are jointly normally distributed. Let  $\Phi_i$  denote the  $2(T_i - 2)$  by  $2(T_i - 2)$  covariance matrix of  $(\eta_{i,f}, \eta_{i,p})$ . In general,  $\Phi_i$  can be written as follows

$$\Phi_i = \begin{pmatrix} \Psi_{i,3} & \Upsilon_{i,3,4} & \cdot & \Upsilon_{i,3,T_i} \\ \Upsilon'_{i,3,4} & \Psi_{i,4} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \Upsilon'_{i,3,T_i} & \cdot & \cdot & \Psi_{i,T_i} \end{pmatrix}, \quad (30)$$

where  $\Psi_{i,t}$  for  $t = 3, \dots, T_i$  denotes the intratemporal covariance matrix of  $(\eta_{i,f,t}, \eta_{i,p,t})$ , and  $\Upsilon_{i,t,s}$  for  $t = 3, \dots, T_i - 1$ ,  $s = 4, \dots, T_i$  denotes the intertemporal covariance matrix of  $(\eta_{i,f,t}, \eta_{i,p,t})$  and  $(\eta_{i,f,s}, \eta_{i,p,s})$ . Attention is restricted to specifications of unobservables in which all parameters occurring in  $\Phi_i$  are identified given only  $\Upsilon_{i,t,s}$  for  $t = 3, \dots, T_i - 1$ ,  $s = 4, \dots, T_i$ .<sup>14</sup>

This requirement restricts attention to distributions of unobservables in which all unobservables with a distribution containing unknown parameters are to some degree persistent. From a behavioral perspective, this requirement has some credentials: it is intuitively plausible that an individual who considers two alternatives to have similar unobserved attributes at one particular point in time will also consider the same two alternatives to have similar unobserved attributes at another point in time.

If  $\max_i\{T_i\} = 4$  then specifications of unobservables with time invariant random coefficients and time invariant random intercepts are permitted. However, all variables with random coefficients must be non-zero for both  $t = 3$  and  $t = 4$  for at least some individuals.

<sup>14</sup>It is further assumed that there is sufficient variation in the data to identify all parameters in the model.

This ensures that the parameters of the covariance matrices of the random coefficients enter both the intra- and intertemporal covariances in observable.<sup>15</sup> If  $\max_i\{T_i\} = 6$  then it is also possible to allow autocorrelated and time invariant components in the random intercepts in addition to time invariant random coefficients on appropriate variables.

Monte Carlo simulations, discussed in Appendix I, illustrate the severity of the fragile identification problem and the empirical properties of the Maximum Likelihood estimator when a specification of unobservables satisfying the above requirement is imposed. The Monte Carlo simulations show that the fragile identification problem is eliminated by imposing appropriate restrictions on the distribution of unobservables. Appendix II presents Monte Carlo evidence demonstrating the desirable empirical properties of the Maximum Simulated Likelihood estimator of the parameters of the two most complex specifications of unobservables considered when estimating the dynamic mixed multinomial logit model of employment dynamics.

## 4 Data and Sample

The data source used for the empirical analysis is the British Household Panel Survey (BHPS). The BHPS is an ongoing panel survey that started in 1991 with a nationally representative sample of approximately 5,500 household in Great Britain. The design of the BHPS is such that the same individuals are re-interviewed in successive years and, if they leave their original households, all adult members of their new households are also interviewed annually. Children in BHPS households are interviewed annually once they reach 16 years of age. Additional samples were added to the main BHPS sample at three different times: in 1997, households in the European Community Household Panel residing in Northern Ireland or in low income households joined; in 1991, 1,500 households in each of Scotland and Wales were added; and in 2001 a sample of 2,000 households was added in Northern Ireland. Households joining from the European Community Household Panel ceased to be members of the BHPS after the 2001 survey, while the other additional households remain

---

<sup>15</sup>Random coefficients are therefore not permitted on time dummies. If random coefficients on a time dummy were included then only employment outcomes in the year corresponding to the time dummy are informative about the parameters of the covariance matrix of the random coefficients on the time dummy. The covariance matrix of the random coefficients on the time dummy will contribute to  $\Psi_{i,t}$  where  $t$  is the year of the time dummy. However, none of the parameters of the covariance matrix of the random coefficients on the time dummy will appear elsewhere in  $\Phi_i$ . Consequently, repeated observations provide no additional information about the covariance matrix of the random coefficients on the time dummy. The fragile identification problem described by Keane (1992) will then occur; parameter estimates are thus likely to be highly unreliable.

in the BHPS.

The sample used for analysis is an unbalanced panel covering the fourteen years 1991-2004. The last year of data for each individual is used purely to construct expectations variables and therefore a maximum of thirteen observations of a woman's employment behavior are available. Attention is restricted to married or cohabiting, non-retired women aged between 18 and 65 years. Single mothers and single person households are therefore excluded from the sample. While resulting in a reduction in the sample size, this feature of the sample selection criteria is desirable as it restricts attention to a relatively homogenous group of women. The sample design is such that a woman's first year in the sample is the first year in which she responded to the BHPS and satisfied the sample criteria. Each woman remains in the sample unless she failed to respond to the BHPS, ceased to be married or cohabiting, retired or reached age 65 years. Furthermore, given that employment transitions are one of the main interests of this study, only women who provide at least four consecutive years of data are included in the sample.

The final sample consists of 4,663 different women. Table 1 shows that number of women joining the sample each year and the distribution of durations in the sample for each cohort of entrants. Less than half of the women in the sample entered at the start of the BHPS in 1991; entry was observed every year with substantial additional numbers of women entering in 1997, 1999 and 2001. While there was a high level of attrition from the sample, a large number of women remained in the sample for six or more years. It should therefore be possible to estimate dynamic mixed multinomial logit models with autocorrelated random intercepts.<sup>16</sup>

Employment outcomes and individual characteristics are measured annually on the anniversary of the date of the first interview with the household. Every year each woman is assigned to either full-time employment, part-time employment or non-employment on the basis of her usual weekly hours of work at the time of the survey. Non-employment corresponds to zero usual weekly hours of work. In accordance with the conventional British definition of part-time work, women reporting usual weekly hours of work of between zero and 30 hours are classified as part-time employed, and women reporting usual weekly hours

---

<sup>16</sup>Due to attrition the women in this sample will not be representative of the corresponding population. However this sample can be used to estimate parameters of interest provided that, conditional on observed characteristics, attrition is unrelated to employment status or, in other words, if there is no selectivity problem. Fitzgerald *et al.* (1998) and Hausman and Wise (1979) provide further discussion of the problems posed by attrition in panel data.

FIRST YEAR IN SAMPLE	NUMBER OF YEARS IN THE SAMPLE												TOTAL
	4	5	6	7	8	9	10	11	12	13	14		
1991	202	126	109	106	101	90	107	87	92	67	808	1895	
1992	14	18	11	10	8	12	6	7	15	54	-	155	
1993	14	11	7	13	3	3	6	5	64	-	-	126	
1994	17	11	6	8	5	3	6	48	-	-	-	104	
1995	9	10	6	9	3	4	60	-	-	-	-	101	
1996	10	9	14	6	5	59	-	-	-	-	-	103	
1997	37	227	13	8	56	-	-	-	-	-	-	341	
1998	34	7	6	72	-	-	-	-	-	-	-	119	
1999	94	75	675	-	-	-	-	-	-	-	-	844	
2000	42	160	-	-	-	-	-	-	-	-	-	202	
2001	673	-	-	-	-	-	-	-	-	-	-	673	
ANY	1146	654	847	232	181	171	185	147	171	121	808	4663	

Table 1: The number of women entering the sample in each year 1991-2001 and the distribution of durations in the sample for each cohort of entrants.

of work of over 30 hours are classified as full-time employed.<sup>17,18</sup> The upper panel of Table 2 shows the percentage of women observed in each employment state for the years 1991-2003. On average, approximately a quarter of women were non-employed, 30% were working part-time and 45% were working full-time. There were no pronounced changes over the sample period in the proportions of women in each employment state.

Figure 1 illustrates the high level of persistence in women's employment outcomes. Figure 1(a) reveals that around 85% of women who were working full-time a year previously are in full-time employment this year. Similarly, Figures 1(b) and 1(c) show that approximately 80% of women who were working part-time or who were non-employed one year previously are in the same employment state this year. There is also evidence of persistence over a longer time horizon. For example, around 55% of women who were working full-time 12 years previously are currently in full-time work. The corresponding figures for part-time work and non-employment are 50% and 39% respectively. It is interesting to note that women who were in part-time employment several years previously are relatively likely to be currently in full-time employment while women who were non-employed a number of years previously are relatively likely to be currently working part-time. These dynamics are consistent with women moving from non-employment into full-time work via a part-time job.

The explanatory variables used in the empirical analysis are the conventional variables used in studies of women's labor supply behavior: measures of education; age; child related variables; and a measure of the women's non-labor incomes. Additionally, measures

<sup>17</sup>Manning and Petrongolo (2005) discuss the relative merits of various definitions of part-time work.

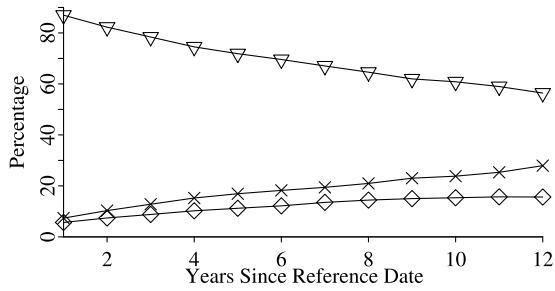
<sup>18</sup>Usual hours of work are taken to be zero for women on maternity leave. Women on maternity leave are therefore classified as non-employed.

VARIABLE	YEAR												
	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2003	
<i>f</i>	43.27	42.00	42.46	43.55	44.38	45.43	45.24	45.82	46.00	45.68	46.16	45.27	44.27
<i>p</i>	29.71	29.37	30.19	29.60	29.23	29.84	29.72	28.70	29.84	29.16	28.44	29.88	31.00
<i>n</i>	27.02	28.63	27.34	26.85	26.39	24.74	25.03	25.48	24.16	25.16	25.40	24.85	24.73
DEGREE	0.11	0.12	0.12	0.12	0.12	0.13	0.13	0.13	0.15	0.15	0.17	0.18	0.18
HIGH QUAL.	0.13	0.14	0.15	0.17	0.17	0.18	0.18	0.20	0.21	0.23	0.23	0.25	0.26
MEDIUM QUAL.	0.09	0.09	0.10	0.10	0.11	0.12	0.12	0.12	0.13	0.13	0.13	0.12	0.12
LOW QUAL.	0.28	0.27	0.27	0.27	0.26	0.25	0.25	0.24	0.24	0.22	0.22	0.21	0.20
INCOME	12.62	12.11	11.89	11.93	11.89	12.08	11.72	11.99	12.42	12.35	12.48	13.06	13.56
AGE	37.80	38.53	38.96	39.08	39.21	39.20	39.25	39.44	39.65	39.84	40.17	40.91	41.82
PRE SCH. CH.	0.28	0.26	0.25	0.24	0.22	0.22	0.22	0.22	0.23	0.24	0.25	0.25	0.24
YOUNGEST CH. 0-1 YEAR	0.07	0.05	0.06	0.06	0.06	0.05	0.06	0.05	0.06	0.07	0.07	0.06	0.05
YOUNGEST CH. 1-2 YEARS	0.11	0.11	0.10	0.09	0.10	0.10	0.10	0.10	0.09	0.09	0.10	0.12	0.10
YOUNGEST CH. 3-4 YEARS	0.06	0.08	0.08	0.08	0.07	0.06	0.08	0.07	0.07	0.07	0.07	0.07	0.08
YOUNGEST CH. 5-6 YEARS	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.06
YOUNGEST CH. 7-11 YEARS	0.12	0.12	0.11	0.12	0.12	0.12	0.12	0.12	0.13	0.14	0.14	0.14	0.13
YOUNGEST CH. 12-15 YEARS	0.10	0.10	0.09	0.10	0.09	0.08	0.09	0.09	0.09	0.09	0.10	0.10	0.11
EXP. CH. IN 0-3 MONTHS	0.02	0.01	0.02	0.02	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.01
EXP. CH. IN 4-6 MONTHS	0.01	0.02	0.01	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
OBSERVATIONS	1895	2050	2176	2078	2039	2001	2197	2178	2885	2905	3193	2962	2729

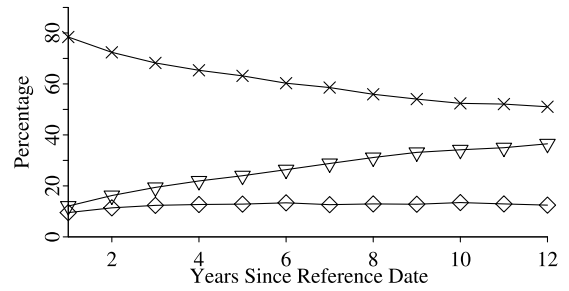
Notes: Calculated from the sample observations for the years 1991-2003. DEGREE is an indicator of a woman having a qualification equivalent to a university degree. HIGH QUAL. is an indicator of the woman's highest qualification being a non-degree level higher qualification such as HND, HNC or nursing qualification. MEDIUM QUAL. is an indicator of a woman's highest qualifications being A levels or equivalent qualifications (A Levels are usually taken at age 18 years). LOW QUAL. is an indicator of the woman's highest qualifications being O levels or grade A-C GCSEs or equivalent qualifications (O levels were replaced by GCSEs, which are usually taken at age 16 years, in 1988). INCOME refers to a woman's annual net non-labor income including her partner's income expressed in pounds sterling and deflated to year 1991 prices using the Retail Price Index. AGE is the woman's age in years. PRE SCH. CH. is the number of children aged under five years in the household. YOUNGEST CH. 0-1 YEAR is an indicator of the youngest child in the household being aged under one year. YOUNGEST CH. 1-2 YEARS, YOUNGEST CH. 3-4 YEARS, YOUNGEST CH. 5-6 YEARS, YOUNGEST CH. 7-11 YEARS and YOUNGEST CH. 12-15 YEARS are similarly defined indicator variables. EXP. CH. IN 0-3 MONTHS and EXP. CH. IN 4-6 MONTHS are indicators of the woman expecting a child in the next three months and in 4-6 months time.

Table 2: Percentage of women in each employment state, sample means of explanatory variables and annual sample sizes.

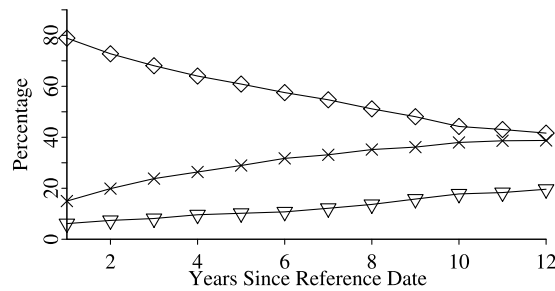




(a) Labor market outcomes of women who were working full-time  $t$  years previously.



(b) Labor market outcomes of women who were working part-time  $t$  years previously.



(c) Labor market outcomes of women who were non-employed  $t$  years previously.

$\nabla$  Full-time                       $\times$  Part-time                       $\diamond$  Non-employment

Figure 1: Observed persistence in employment outcomes.

of short-term fertility expectations are constructed in order to capture any adjustments in women's labor supply behavior shortly before the arrival of a child. The chosen measures of educational qualifications correspond to the main levels of educational attainment in the UK, while the child variables are sufficiently fine so as to capture variation in the availability and cost of child-care for children of different ages. Non-labor income is defined as household income, excluding any income the woman obtains from employment. Furthermore, tax paid on investment income and payroll deductions, including income tax and any pension contributions, are also deducted when calculating non-labor income. Further details concerning the explanatory variables are provided in the notes accompanying Table 2.

The lower panel of Table 2 details the sample means of the explanatory variables. The average age of the sampled women increased by approximately four years between 1991 and 2003. The proportions of women whose highest qualification was a degree level qualification or a non-degree level higher qualification increased over the sample period, as did the proportion of women whose highest qualifications were A levels. These increases were balanced by decreases in the proportion of women whose highest qualifications were GCSEs and the proportion of women with no qualifications. The rising levels of educational qualifications

EMPLOYMENT STATE	QUALIFICATIONS				INCOME			
	NONE	LOW	MED.	HIGH	DEGREE	LOW	HIGH	
<i>f</i>	51.27	43.48	53.45	54.84	57.68	45.93	32.28	
<i>p</i>	28.38	30.92	25.84	28.39	26.16	28.80	37.64	
<i>n</i>	20.35	25.60	20.71	16.76	16.15	25.27	30.07	
	YOUNGEST CHILD			EXPECTING CHILD				
	NONE	<2 YEARS	2≤YEARS<6	YES		NO		
<i>f</i>	29.98	20.43	25.49	33.25		45.88		
<i>p</i>	36.22	30.08	40.09	14.21		39.78		
<i>n</i>	33.80	49.49	34.42	52.53		35.34		

Notes: Calculated from the sample observations for the years 1991-2003. NO QUAL. refers to women with no qualifications or qualifications below GCSE level. LOW INCOME refers to women with below mean non-labor incomes and HIGH INCOME refers to women with above mean non-labor incomes. All other variables are as defined in the notes accompanying Table 2.

Table 3: Percentages of women in full-time work, part-time work and non-employment according to individual characteristics.

was due primarily to the replacement of old sample members with new sample members with higher average levels of education; only a small number of sample members acquired qualifications during their time in the sample. Non-labor income displayed an upwards trend over the sample period, while there were no pronounced trends in the sample means of the child related variables.

Table 3 shows the employment outcomes of the sampled women according to various individual characteristics. The figures presented in Table 3 confirm many of the established features of women's labor supply behavior. In particular, the higher a woman's level of qualifications the more likely she is to be working full-time and the less likely she is to be non-employed. Women whose youngest child is aged under two years have a 20% higher probability of non-employment as compared to women without children, and are also less likely to be working part-time as compared to childless women. Women whose youngest child is aged between two and four years are slightly more likely than childless women to be non-employed. At the intensive margin, women whose youngest child is aged between two and four years are less likely to be working full-time and more likely to be working part-time as compared to childless women. The income effect is in the expected direction: women with above average non-labor incomes are less likely to be working full-time and are more likely to be non-employed than women with below average non-labor incomes.

This analysis is purely descriptive. The results of the more formal analysis are of far greater interest as they are informative about the causal effects of past employment outcomes and observed individual characteristics on women's labor supply choices.

## 5 Results I: Dynamic Mixed Multinomial Logit Models

The dynamic mixed multinomial logit model is estimated with six different specifications of the unobservables  $\eta_{i,f}$  and  $\eta_{i,p}$ . All six specifications are nested within the structure described above in Section 3.1.2 and satisfy the requirement discussed in Section 3.3 which ensure that the fragile identification problem is avoided. The results from the specification with the most general distribution of unobservables are discussed in detail. Comparisons are made with the results from specifications with more restrictive distributions of unobservables. This exercise allows an exploration of the implications of different assumptions regarding the distribution of unobservables for conclusions concerning state dependencies in labor supply behavior.

Specification VI, the most general specification of the dynamic mixed multinomial logit model under consideration, allows random intercepts with both time invariant and auto-correlated components and time invariant random coefficients on the indicators of having a degree and the woman's youngest child being aged under one year. Time invariant random coefficients on the lagged dependent variables and the initial conditions are also included.<sup>19</sup> The last two columns in Table 4 show the deterministic components of the coefficients on the lagged dependent variables and explanatory variables. The coefficient estimates are as expected and are not discussed. Instead attention is focused on the marginal effects reported in Table 7. For a woman with no qualifications, any increase in the level of her qualifications significantly increases her probability of working full-time, but has no significant effect on her probability of working part-time. There is a small but significant negative income effect for full-time work; a £1000 per year increase in a woman's non-labor income decreases her probability of working full-time by 0.08(0.01) percentage points. In contrast, changes in non-labor income do not significantly affect the woman's probability of working part-time. Young children have a very strong negative effect on a woman's likelihood of working full-time. Specifically, women whose youngest child is aged under one year are 33.87(1.82) percentage points less likely to be working full-time than women without children. The effect of children on a woman's probability of engaging in full-time work decreases quickly as the age of the woman's youngest child increases. Indeed, a woman whose youngest child is aged between 12 and 16 years has the same probability of working full-time as an otherwise identical woman without children. A youngest child aged between 1 and 7 years has a large

<sup>19</sup>Experimentation with various specifications of the random coefficients revealed that there are no random coefficients with significant amounts of variations on any of the other explanatory variables.

positive effect on a woman's probability of working part-time. Women who are expecting a child in the next 6 months are significantly less likely to work full-time than women not expecting a child, while women expecting a child in the next three month have a significantly lower probability of working part-time than women not expecting a child.

Even after allowing unobservables with a relatively general intra- and intertemporal distribution, significant own state dependencies in both full-time and part-time work are evident. Specifically, the marginal effects for Specification VI shown in Table 7 reveal that working full-time rather than being non-employed in the previous year increases the probability of working full-time in the current year by 44.24(2.43) percentage points. Similarly, working part-time rather than being non-employed increases a woman's probability of working part-time in the current year by 24.62(2.41) percentage points.

An alternative method of exploring the structure of state dependencies in women's labor supply behavior is to simulate the effects of labor market policies that temporally affect women's incentives to work full-time or to work part-time. Two artificial policy interventions are considered. The design of the policies is such that they cause women who chose non-employment in their second year in the sample to move into, depending on the policy, either full-time or part-time work. In the absence of any state dependencies in labor supply these polices will affect labor market outcomes only for the duration of the policy. In contrast, in the presence of state dependencies the policies will affect labor supply outcomes over a longer period of time. Figure 2 shows the long-term effects of these policies on the rates of full-time and part-time employment amongst women affected by the policies.

The results for Specification VI imply that one year after a policy incentivizing non-employed women to move into full-time work, the rate of full-time employment amongst women affected by the policy is 30.92(2.74) percentage points higher than in the absence of the policy. The effect of the policy diminishes over time, reaching just 3 percentage points a decade after the policy. Similarly, a policy incentivizing non-employed women to move into part-time work leads to a 30.99(2.89) percentage point increase in the rate of part-time employment amongst women affected by the policy. Again the effect of the policy decreases over time, and is negligible ten years after the policy is withdrawn. Standard errors (not reported) indicate that the former policy leads to a significant increase in full-time employment in each of the eleven years following the policy, while the latter policy produces a significant increase in the rate of part-time employment for nine years following the policy. These results confirm the presence of state dependencies in labor supply that

remain after controls for persistent unobserved individual characteristics.

In terms of the distribution of unobservables, the results in Table 10 show negative first order autocorrelation in the unobservables affecting payoffs from full-time employment and positive first order autocorrelation in the unobservables affecting payoffs from part-time employment. Women with young children have very large amounts of variation in their unobserved payoffs from working full-time, possibly reflecting variation in child-care costs or productivity in home production. There is also significant variation in women's unobserved preferences for part-time employment if they have a young child, but far less than for full-time employment. Similarly, women with a degree level qualification have a significantly higher level of unobserved variation in their payoffs from working full-time than women with other levels of qualifications.

There are some difference in the results implied by Specification VI and the results obtained when more restrictive distributions of unobservables are imposed. The most restrictive distribution of unobservables under consideration is Specification I. In this specification, random components in  $\eta_{i,f}$  and  $\eta_{i,p}$  are excluded and thus there is no persistence in the random component of unobservables. However, non-random coefficients on the initial conditions are permitted. Comparing the marginal effects reported in Table 7 across the two specifications shows that the own state dependencies in employment behavior are larger in magnitude when persistent unobservables are excluded. Figure 2 also shows that Specification I implies substantially larger own state and cross state dependencies following interventions in the labor market; as expected, ignoring persistent unobservables leads to an overestimate of the state dependencies in labor supply behavior. In contrast, the effects of education, income, children and fertility expectations are broadly comparable across the two specifications and indeed the other specifications of unobservables under consideration.

VARIABLE	SPEC. I		SPEC. II		SPEC. III		SPEC. IV		SPEC. V		SPEC. VI	
	f	p	f	p	f	p	f	p	f	p	f	p
$Y_{i,f,t-2}$	2.19 (0.11)	1.04 (0.09)	1.59 (0.13)	0.66 (0.11)	1.59 (0.13)	0.67 (0.11)	1.98 (0.26)	0.82 (0.16)	1.72 (0.18)	0.78 (0.16)	2.13 (0.29)	0.83 (0.18)
$Y_{i,p,t-2}$	1.09 (0.11)	1.37 (0.07)	0.72 (0.13)	0.95 (0.09)	0.72 (0.13)	0.95 (0.09)	0.88 (0.17)	1.02 (0.12)	0.77 (0.16)	1.01 (0.12)	0.92 (0.15)	1.14 (0.15)
$Y_{i,f,t-1}$	3.81 (0.10)	1.50 (0.09)	2.90 (0.13)	0.98 (0.11)	2.93 (0.13)	1.01 (0.11)	4.27 (0.46)	1.01 (0.21)	3.50 (0.22)	1.16 (0.16)	5.28 (0.57)	1.37 (0.27)
$Y_{i,p,t-1}$	2.09 (0.11)	2.85 (0.07)	1.55 (0.13)	2.34 (0.09)	1.55 (0.13)	2.34 (0.09)	1.87 (0.26)	2.38 (0.15)	1.83 (0.19)	2.43 (0.10)	2.28 (0.31)	2.52 (0.16)
DEGREE	0.66 (0.10)	0.30 (0.09)	1.00 (0.16)	0.54 (0.13)	0.98 (0.16)	0.53 (0.13)	1.23 (0.21)	0.63 (0.15)	1.04 (0.17)	0.60 (0.14)	1.24 (0.16)	0.66 (0.16)
HIGH QUAL.	0.67 (0.09)	0.34 (0.08)	1.00 (0.13)	0.57 (0.11)	0.98 (0.13)	0.55 (0.11)	1.24 (0.18)	0.66 (0.12)	1.03 (0.14)	0.64 (0.12)	1.24 (0.19)	0.72 (0.14)
MEDIUM QUAL.	0.41 (0.11)	0.12 (0.09)	0.62 (0.16)	0.26 (0.14)	0.62 (0.16)	0.26 (0.14)	0.77 (0.21)	0.32 (0.15)	0.59 (0.17)	0.27 (0.15)	0.73 (0.21)	0.31 (0.16)
LOW QUAL.	0.34 (0.08)	0.14 (0.07)	0.50 (0.13)	0.25 (0.10)	0.48 (0.13)	0.24 (0.10)	0.63 (0.16)	0.30 (0.11)	0.58 (0.15)	0.35 (0.13)	0.69 (0.18)	0.41 (0.14)
INCOME	-0.01 (0.00)	0.00 (0.00)	-0.01 (0.00)	0.00 (0.00)	-0.01 (0.00)	0.00 (0.00)	-0.02 (0.00)	0.00 (0.00)	-0.01 (0.00)	0.00 (0.00)	-0.02 (0.00)	-0.01 (0.00)
AGE/10	0.42 (0.25)	0.34 (0.24)	0.69 (0.35)	0.43 (0.31)	0.54 (0.35)	0.30 (0.31)	0.99 (0.43)	0.57 (0.34)	0.46 (0.38)	0.33 (0.34)	0.31 (0.46)	0.28 (0.37)
AGE <sup>2</sup> /100	-0.09 (0.03)	-0.04 (0.03)	-0.14 (0.04)	-0.05 (0.04)	-0.12 (0.04)	-0.03 (0.04)	-0.19 (0.05)	-0.07 (0.04)	-0.11 (0.05)	-0.04 (0.04)	-0.11 (0.06)	-0.04 (0.05)
PRE SCH. CH.	0.45 (0.12)	0.13 (0.09)	0.23 (0.15)	0.06 (0.11)	0.23 (0.15)	0.06 (0.11)	0.35 (0.19)	0.07 (0.12)	0.01 (0.17)	-0.03 (0.12)	0.09 (0.22)	0.01 (0.13)
YOUNGEST CH. 0-1 YEAR	-4.46 (0.19)	-1.57 (0.17)	-5.19 (0.24)	-2.00 (0.20)	-5.12 (0.24)	-1.96 (0.20)	-6.73 (0.63)	-2.38 (0.31)	-5.80 (0.39)	-1.92 (0.26)	-7.79 (0.85)	-2.31 (0.35)
YOUNGEST CH. 1-2 YEARS	-1.30 (0.17)	0.41 (0.15)	-2.07 (0.22)	0.05 (0.18)	-1.99 (0.22)	0.10 (0.18)	-2.72 (0.38)	-0.13 (0.21)	-1.73 (0.25)	0.14 (0.20)	-2.27 (0.39)	-0.06 (0.23)
YOUNGEST CH. 3-4 YEARS	-0.91 (0.15)	0.34 (0.12)	-1.49 (0.19)	0.15 (0.15)	-1.42 (0.19)	0.21 (0.15)	-1.98 (0.30)	0.06 (0.17)	-1.27 (0.20)	0.23 (0.16)	-1.68 (0.31)	0.11 (0.18)
YOUNGEST CH. 5-6 YEARS	-0.23 (0.14)	0.56 (0.12)	-0.69 (0.18)	0.49 (0.14)	-0.63 (0.18)	0.54 (0.15)	-0.98 (0.25)	0.44 (0.16)	-0.57 (0.19)	0.56 (0.16)	-0.75 (0.26)	0.54 (0.17)
YOUNGEST CH. 7-11 YEARS	-0.10 (0.11)	0.35 (0.10)	-0.29 (0.15)	0.41 (0.13)	-0.23 (0.14)	0.46 (0.13)	-0.46 (0.19)	0.40 (0.14)	-0.19 (0.15)	0.47 (0.14)	-0.28 (0.20)	0.48 (0.15)
YOUNGEST CH. 12-15 YEAR	0.30 (0.13)	0.27 (0.12)	0.27 (0.16)	0.32 (0.15)	0.33 (0.16)	0.36 (0.15)	0.28 (0.19)	0.34 (0.16)	0.34 (0.17)	0.37 (0.16)	0.39 (0.20)	0.37 (0.17)
EXP. CH. IN 0-3 MONTHS	-1.44 (0.19)	-1.14 (0.21)	-1.81 (0.22)	-1.39 (0.22)	-1.82 (0.22)	-1.39 (0.22)	-2.32 (0.33)	-1.58 (0.24)	-2.31 (0.26)	-1.68 (0.23)	-2.96 (0.41)	-1.83 (0.26)
EXP. CH. IN 4-6 MONTHS	-0.61 (0.26)	-0.38 (0.20)	-0.86 (0.31)	-0.49 (0.22)	-0.86 (0.31)	-0.50 (0.22)	-1.08 (0.40)	-0.53 (0.23)	-0.96 (0.33)	-0.55 (0.24)	-1.23 (0.43)	-0.61 (0.25)
INTERCEPT	-3.59 (0.48)	-3.20 (0.47)	-4.11 (0.68)	-3.41 (0.63)	-4.61 (0.79)	-3.66 (0.78)	-5.69 (0.95)	-3.72 (0.69)	-4.10 (0.78)	-3.28 (0.69)	-4.97 (1.05)	-3.22 (0.75)

Notes: Standard errors in parenthesis. Estimates of the coefficients on time dummies and the initial conditions are omitted.

Table 4: Estimates of the deterministic components of the coefficients for Specifications I-VI of the dynamic mixed multinomial logit model.

	SPEC. II	SPEC. III	SPEC. IV	SPEC. V	SPEC. VI
$\Sigma_{Intercept\ 1}$	$\begin{pmatrix} 2.78 & . \\ (0.32) & . \end{pmatrix}$	$\begin{pmatrix} 1.52 & . \\ (0.62) & . \end{pmatrix}$	$\begin{pmatrix} 4.37 & . \\ (0.74) & . \end{pmatrix}$	$\begin{pmatrix} 0.18 & . \\ (0.39) & . \end{pmatrix}$	$\begin{pmatrix} 0.09 & . \\ (0.25) & . \end{pmatrix}$
$\Sigma_{Intercept\ 2}$		$\begin{pmatrix} 1.00 & 1.18 \\ (0.98) & (3.13) \end{pmatrix}$	$\begin{pmatrix} 2.24 & 1.70 \\ (0.38) & (0.30) \end{pmatrix}$	$\begin{pmatrix} 0.32 & 0.62 \\ (0.40) & (0.36) \end{pmatrix}$	$\begin{pmatrix} 0.20 & 0.58 \\ (0.35) & (0.32) \end{pmatrix}$
$\mu_2$		$\begin{pmatrix} 3.30 & . \\ (1.08) & . \end{pmatrix}$			
$\alpha$		$\begin{pmatrix} 1.48 & 0.83 \\ (0.56) & (0.38) \end{pmatrix}$			
$\rho_f$			$\begin{pmatrix} 2.08 \\ (2.13) \end{pmatrix}$		
$\rho_p$			$\begin{pmatrix} 1.47 \\ (1.10) \end{pmatrix}$		
$\Sigma_{\zeta}$				$\begin{pmatrix} -0.17 \\ (0.10) \end{pmatrix}$	$\begin{pmatrix} -0.12 \\ (0.10) \end{pmatrix}$
$\Sigma_{Y_{i,f,t-2}}$				$\begin{pmatrix} 0.45 \\ (0.29) \end{pmatrix}$	$\begin{pmatrix} 0.43 \\ (0.26) \end{pmatrix}$
$\Sigma_{Y_{i,p,t-2}}$			$\begin{pmatrix} 3.35 & . \\ (1.33) & . \end{pmatrix}$		$\begin{pmatrix} 4.08 & . \\ (1.59) & . \end{pmatrix}$
$\Sigma_{Y_{i,f,t-1}}$			$\begin{pmatrix} 0.93 & 0.38 \\ (0.35) & (0.32) \end{pmatrix}$		$\begin{pmatrix} 1.18 & 0.47 \\ (0.53) & (0.37) \end{pmatrix}$
$\Sigma_{Y_{i,p,t-1}}$				$\begin{pmatrix} 0.93 & . \\ (0.43) & . \end{pmatrix}$	$\begin{pmatrix} 0.00 & . \\ (0.00) & . \end{pmatrix}$
$\Sigma_{DEGREE_{i,t}}$				$\begin{pmatrix} 0.61 & 0.43 \\ (0.34) & (0.40) \end{pmatrix}$	$\begin{pmatrix} 0.00 & 0.17 \\ (0.16) & (0.29) \end{pmatrix}$
$\Sigma_{CHILD<1_{i,t}}$				$\begin{pmatrix} 0.74 & . \\ (0.48) & . \end{pmatrix}$	$\begin{pmatrix} 1.32 & . \\ (0.76) & . \end{pmatrix}$
				$\begin{pmatrix} 0.63 & 0.56 \\ (0.32) & (0.28) \end{pmatrix}$	$\begin{pmatrix} 0.93 & 0.66 \\ (0.43) & (0.33) \end{pmatrix}$
				$\begin{pmatrix} 0.36 & . \\ (0.38) & . \end{pmatrix}$	$\begin{pmatrix} 0.75 & . \\ (0.58) & . \end{pmatrix}$
				$\begin{pmatrix} 0.35 & 0.71 \\ (0.38) & (0.55) \end{pmatrix}$	$\begin{pmatrix} 0.51 & 0.58 \\ (0.45) & (0.55) \end{pmatrix}$
				$\begin{pmatrix} 0.77 & . \\ (0.61) & . \end{pmatrix}$	$\begin{pmatrix} 0.92 & . \\ (0.99) & . \end{pmatrix}$
				$\begin{pmatrix} 0.26 & 0.09 \\ (0.27) & (0.16) \end{pmatrix}$	$\begin{pmatrix} 0.39 & 0.17 \\ (0.37) & (0.25) \end{pmatrix}$
				$\begin{pmatrix} 1.02 & . \\ (0.56) & . \end{pmatrix}$	$\begin{pmatrix} 1.96 & . \\ (0.91) & . \end{pmatrix}$
				$\begin{pmatrix} 0.61 & 0.48 \\ (0.39) & (0.35) \end{pmatrix}$	$\begin{pmatrix} 0.97 & 0.52 \\ (0.54) & (0.40) \end{pmatrix}$
				$\begin{pmatrix} 14.31 & . \\ (3.04) & . \end{pmatrix}$	$\begin{pmatrix} 27.15 & . \\ (7.22) & . \end{pmatrix}$
				$\begin{pmatrix} 6.41 & 3.03 \\ (1.57) & (0.96) \end{pmatrix}$	$\begin{pmatrix} 10.33 & 3.97 \\ (2.59) & (1.29) \end{pmatrix}$

Notes: Standard errors in parenthesis. Specification I has no unknown parameters in the distribution of unobservables. In Specifications II and IV-VI,  $\Sigma_{Intercept\ 1}$  is the covariance matrix of the time invariant components of the random intercepts. Specification III has time invariant random intercepts with a distribution obtained from the mixture of two bivariate normal distributions: with probability  $\alpha$  the random intercepts have mean zero and variance  $\Sigma_{Intercept\ 1}$  and with probability  $(1 - \alpha)$  the random intercepts have mean  $\mu_2$  and variance  $\Sigma_{Intercept\ 2}$ . In specifications allowing autocorrelation in the random intercepts,  $\rho_f$  and  $\rho_p$  are the first order autocorrelation coefficients and  $\Sigma_{\zeta}$  is the covariance matrix of the innovations in the autoregressive processes.  $\Sigma_{DEGREE_{i,t}}$  and  $\Sigma_{CHILD<1_{i,t}}$  are the covariance matrices of the random coefficients on the indicated variables. The covariance matrices of the random coefficients on the initial conditions in Specifications V and VI are not reported.

Table 5: Estimates of parameters appearing in the distribution of unobservables for Specifications II-VI of the dynamic mixed multinomial logit model.

	SPEC. I	SPEC. II	SPEC. III	SPEC. IV	SPEC. V	SPEC. VI
Log likelihood	-17524.43	-10697.87	-10690.95	-10683.49	-10622.91	-10607.21
AIC	36849.94	21557.75	21555.90	21538.98	21473.82	21452.43
BIC	36885.00	22205.51	22251.64	22226.73	22385.49	22404.08
Pseudo R <sup>2</sup>	47.57%	54.57%	54.60%	54.63%	54.88%	54.95%
Notes:	AIC=-2Log likelihood+2Parameters;		BIC=-2Log likelihood+		ln(NT)Parameters; Pseudo R <sup>2</sup> =1-Restricted Log likelihood/Unrestrcited Log likelihood.	

Table 6: Maximized log likelihood values, model selection criteria and pseudo R<sup>2</sup> values for Specifications I-VI of the dynamic mixed multinomial logit model.

Specification II generalizes Specification I by allowing the two intercepts to have time invariant random components. While still substantially more restrictive than Specification VI, this specification allows persistence in unobservables and intratemporal heteroscedasticity and correlations in unobservables across employment states. In terms of dynamic responses to policy interventions, Specification II implies lower own state dependencies in both full-time and part-time employment than those implied by either Specification I or Specification IV; it is not unambiguously the case that more flexible distributions of unobservables lead to lower estimates of state dependencies in labor supply behavior. Specification III allows time invariant random intercepts with a distribution formed from a mixture of two bivariate normal distributions. This specification is included as mixtures of normals have previously been found to provide a superior representation of unobserved heterogeneity (see Geweke *et al.*, 1997). However, while the parameter estimates in Table 10 show significant differences in the means and variances of the two components of the mixture distribution, the marginal effects and dynamic responses differ very little between Specifications II and III.

Specification IV allows random intercepts with both time invariant and autocorrelated components but random coefficients are excluded. This generalization leads to changes in the implied nature of state dependencies in labor supply behavior: state dependence in full-time employment is higher according to Specification IV than according to Specification II, although the implied level of state dependence in part-time employment is robust to this generalization. Lastly, Specification V excludes autocorrelation in the random intercepts but allows random coefficients as described above. Interestingly, this specification of unobservables generates predictions similar to those produced by Specification IV.

Given the additional complexity encountered when extending the distribution of unobservables from time invariant random intercepts to allow autocorrelation or random coefficients, it is important to determine whether specifications allowing additional flexibility in the distribution of unobservables dominate more restrictive specifications. Model selec-



VARIABLE	SPEC. I		SPEC. II		SPEC. III		SPEC. IV		SPEC. V		SPEC. VI	
	f	p	f	p	f	p	f	p	f	p	f	p
$n_{t-1} \rightarrow f_{t-1}$	50.80 (1.63)	-7.79 (1.39)	31.37 (1.71)	-9.45 (1.71)	31.48 (3.54)	-8.98 (3.02)	37.95 (2.48)	-15.61 (2.05)	36.00 (2.72)	-11.76 (2.89)	44.24 (2.43)	-18.03 (2.38)
$n_{t-1} \rightarrow p_{t-1}$	3.47 (1.09)	39.97 (1.31)	-2.37 (1.95)	27.77 (2.27)	-1.88 (1.97)	27.61 (1.75)	-0.22 (1.96)	25.02 (1.94)	-0.61 (1.94)	26.93 (3.47)	2.04 (1.71)	24.62 (2.41)
NO EDUC. $\rightarrow$ LOW EDUC.	3.83 (0.67)	-0.40 (0.80)	5.81 (1.14)	0.11 (1.09)	5.67 (0.94)	-0.05 (1.07)	5.50 (1.14)	0.28 (1.20)	5.29 (0.98)	0.45 (0.92)	4.62 (1.06)	0.88 (0.98)
NO EDUC. $\rightarrow$ MED. EDUC.	3.66 (0.59)	-0.06 (0.70)	5.63 (0.91)	0.41 (0.97)	5.53 (0.93)	0.20 (1.19)	5.38 (0.82)	0.56 (0.83)	4.98 (0.90)	0.94 (0.88)	4.36 (0.87)	1.43 (0.88)
NO EDUC. $\rightarrow$ HIGH EDUC.	2.70 (0.61)	-0.91 (0.75)	4.02 (0.99)	-0.67 (0.91)	4.11 (1.09)	-0.82 (1.30)	3.75 (1.07)	-0.41 (0.89)	3.49 (0.98)	-0.53 (1.20)	3.09 (0.95)	-0.10 (1.16)
NO EDUC. $\rightarrow$ DEGREE	2.05 (0.48)	-0.33 (0.62)	3.04 (0.99)	-0.10 (0.84)	2.92 (0.93)	-0.13 (0.96)	2.88 (0.76)	0.09 (0.84)	2.83 (0.71)	0.21 (0.63)	2.49 (0.68)	0.44 (0.86)
INCOME+£1000	-0.08 (0.02)	0.03 (0.02)	-0.08 (0.02)	0.04 (0.02)	-0.10 (0.02)	0.03 (0.02)	-0.11 (0.02)	0.05 (0.02)	-0.10 (0.02)	0.03 (0.02)	-0.08 (0.02)	0.03 (0.02)
AGE + 1 YEAR	-0.28 (0.04)	0.22 (0.04)	-0.42 (0.04)	0.27 (0.05)	-0.44 (0.05)	0.29 (0.05)	-0.40 (0.05)	0.27 (0.06)	-0.40 (0.04)	0.25 (0.04)	-0.36 (0.04)	0.23 (0.05)
PRE. SCH. CH. + 1	3.08 (0.98)	-1.26 (0.92)	1.73 (1.18)	-0.83 (0.96)	1.67 (1.27)	-0.74 (1.16)	2.05 (1.04)	-1.05 (1.15)	0.22 (1.17)	-0.35 (1.05)	0.52 (1.13)	-0.34 (1.08)
NO CHILDREN $\rightarrow$ CHILD AGED $\leq$ 1 YEAR	-36.73 (1.72)	2.31 (1.94)	-38.15 (1.61)	1.99 (2.31)	-37.99 (4.29)	1.84 (3.18)	-37.25 (1.47)	0.96 (2.11)	-34.61 (1.78)	1.30 (2.34)	-33.87 (1.82)	-0.21 (2.46)
NO CHILDREN $\rightarrow$ 1 YEAR <CHILD AGED $\leq$ 2 YEARS	-15.61 (1.77)	14.08 (1.54)	-21.51 (1.80)	15.58 (2.03)	-20.93 (2.67)	15.52 (2.55)	-20.23 (1.77)	13.98 (2.00)	-17.83 (1.91)	13.96 (2.03)	-15.83 (2.05)	11.78 (2.08)
NO CHILDREN $\rightarrow$ 2 YEAR <CHILD AGED $\leq$ 4 YEARS	-11.19 (1.27)	10.37 (1.22)	-16.44 (1.63)	12.98 (1.70)	-16.13 (1.97)	13.11 (1.98)	-15.77 (1.45)	12.18 (1.70)	-14.14 (1.54)	12.00 (1.66)	-12.77 (1.51)	10.70 (1.75)
NO CHILDREN $\rightarrow$ 4 YEAR <CHILD AGED $\leq$ 7 YEARS	-6.33 (1.10)	8.49 (0.99)	-11.04 (1.26)	11.71 (1.40)	-10.51 (1.49)	11.61 (1.51)	-10.52 (1.30)	11.03 (1.39)	-9.77 (1.39)	11.24 (1.44)	-8.68 (1.12)	10.26 (1.29)
NO CHILDREN $\rightarrow$ 7 YEAR <CHILD AGED $\leq$ 11 YEARS	-3.78 (0.82)	5.25 (0.88)	-6.41 (1.07)	7.69 (1.23)	-6.04 (0.98)	7.68 (1.06)	-6.37 (0.96)	7.53 (0.97)	-5.57 (0.82)	7.42 (0.95)	-5.11 (0.99)	7.13 (1.05)
NO CHILDREN $\rightarrow$ 11 YEAR <CHILD AGED $\leq$ 16 YEARS	0.34 (1.47)	1.89 (1.04)	-0.09 (1.03)	2.53 (1.09)	0.22 (1.23)	2.56 (1.10)	-0.37 (0.92)	2.58 (1.09)	0.15 (0.90)	2.62 (1.02)	0.08 (1.11)	2.47 (1.10)
NOT EXP. CHILD $\rightarrow$ EXP. CHILD IN 0-3 MONTHS	-6.42 (1.54)	-5.37 (2.12)	-8.72 (1.75)	-6.05 (2.30)	-8.96 (2.02)	-6.09 (2.60)	-8.74 (1.82)	-6.41 (1.97)	-10.61 (1.83)	-6.89 (2.38)	-10.59 (2.10)	-6.43 (2.81)
NOT EXP. CHILD $\rightarrow$ EXP. CHILD IN 4-6 MONTHS	-2.91 (2.04)	-0.94 (1.93)	-4.81 (2.36)	-0.49 (2.07)	-4.79 (2.29)	-0.65 (2.60)	-4.68 (2.17)	-0.53 (1.94)	-4.85 (2.14)	-0.82 ( )	-4.78 (2.46)	-0.55 (2.20)

Notes: Standard errors in parenthesis. Marginal effects are expressed in percentages and are averages over the sampled individuals and over time.

Table 7: Marginal effects for Specifications I-VI of the dynamic mixed multinomial logit model.

tion criteria presented in Table 6 are inconclusive regarding the preferred specification of unobservables: the Akaike Information Criterion (Akaike, 1973) suggests Specification VI is preferred, while the Bayesian Information Criterion (Schwartz, 1978) selects Specification II. Goodness of fit analysis, presented in Appendix II shows that all six specifications of the dynamic mixed multinomial logit model accurately predict the observed percentage of women in each employment state in each year (see Table 15). Furthermore Table 16 shows that all six specifications provide reasonable predictions of the nature of persistence in employment outcomes. However, Specifications IV and VI, which both include autocorrelated random intercepts, provide more accurate predictions than the other specifications.

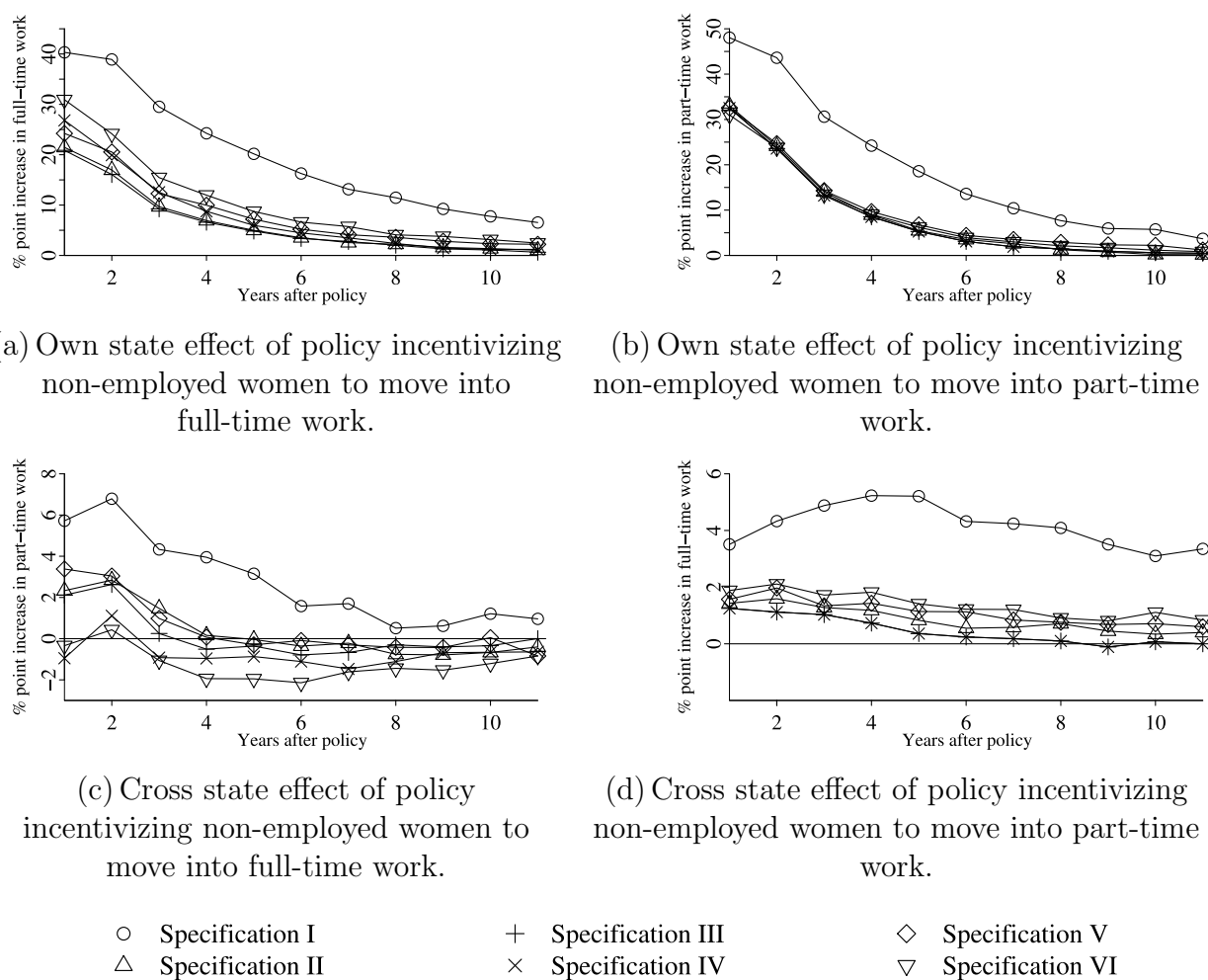


Figure 2: Dynamic effects of policy interventions for Specifications I-VI of the dynamic mixed multinomial logit model.

More information is gleaned from comparing the dynamic responses to temporary policy interventions implied by the different specifications. Table 8 shows t tests for the significance of the differences in the dynamic responses shown in Figure 2. The tests are conducted under the null hypothesis that Specification II is correct. There are clearly significant differences

in the dynamic responses implied by Specification II and those implied by the more general specifications. In many years and for both full-time and part-time employment the dynamic responses are significantly biased downwards by working with a specification of unobservables that allows time invariant random intercepts but no further generality in the distribution of unobservables. However, while the different specifications imply significantly different rates of part-time employment following a policy incentivizing non-employed women to move into part-time work, Figure 2(b) shows that the magnitudes of these differences are small. The differences in the rates of part-time employment are therefore unlikely to be economically important. In contrast, the differences between the own state effects for full-time work are large enough to make the choice of distribution of unobservables important when conducting policy evaluation.

EMPLOYMENT STATE	YEARS SINCE POLICY INTERVENTION										
	1	2	3	4	5	6	7	8	9	10	11
DIFFERENCE BETWEEN SPEC. II AND SPEC. IV											
<i>f</i>	23.45	6.44	6.30	3.77	1.94	2.62	3.41	0.38	-0.42	-0.56	0.52
<i>p</i>	-6.71	-1.87	-3.87	-2.78	-1.51	-2.32	-2.78	-0.28	1.14	1.98	1.63
DIFFERENCE BETWEEN SPEC. II AND SPEC. V											
<i>f</i>	1.43	2.15	1.90	2.92	2.64	2.23	2.52	2.52	1.99	1.98	3.25
<i>p</i>	-0.51	0.34	0.31	0.97	2.18	1.63	2.16	3.50	4.75	6.47	2.94
DIFFERENCE BETWEEN SPEC. II AND SPEC. VI											
<i>f</i>	2.97	2.42	2.79	2.56	2.59	2.17	2.64	1.65	1.94	1.73	1.69
<i>p</i>	-1.24	-0.16	-0.40	0.12	0.79	0.47	0.72	1.67	2.94	3.01	2.17

Notes: Significant tests are conducted under the null that Specification II is the true specification. Standard errors are bootstrapped.

Table 8: t tests for significance of differences in the own state effects of the two policy interventions.

## 6 Results II: Dynamic Linear Probability Models

Table 9 reports estimates of the parameter of the dynamic linear probability models of full-time and part-time work obtained from Generalized Method of Moments estimation of the first difference Equations (28) and (29) and Ordinary Least Squares estimation of the corresponding level equations. The preferred results are those obtained from Generalized Method of Moments estimation as this estimation method allows time invariant individual specific fixed effects. Where appropriate, the Generalized Method of Moments results are compared to the results from the dynamic mixed multinomial logit model discussed above.

Instruments for the difference dynamic linear probability model of full-time work consist of employment outcomes, income and child variables in the years  $t - 2$ ,  $t - 3$  and  $t - 4$ , and

VARIABLE	FULL-TIME WORK		PART-TIME WORK	
	OLS LEVELS	DIFF. GMM	OLS LEVELS	DIFF. GMM
$Y_{i,f,t-1}$	56.91 (1.01)	35.97 (2.89)	-5.22 (1.15)	-7.21 (2.81)
$Y_{i,p,t-1}$	2.19 (0.74)	-3.83 (1.64)	50.05 (1.01)	33.91 (2.47)
$Y_{i,f,t-2}$	24.97 (0.98)	9.03 (1.53)	1.85 (1.13)	-0.44 (1.80)
$Y_{i,p,t-2}$	1.99 (0.74)	-1.15 (1.11)	22.46 (1.04)	8.17 (1.69)
DEGREE	4.54 (0.68)	-	-0.26 (0.75)	-
HIGH QUAL.	4.33 (0.61)	-	0.29 (0.68)	-
MEDIUM QUAL.	3.40 (0.72)	-	-0.51 (0.80)	-
LOW QUAL.	2.51 (0.58)	-	0.00 (0.67)	-
PRE SCH. CH.	6.51 (0.93)	-0.19 (1.04)	-2.41 (1.18)	-0.81 (1.59)
YOUNGEST CH. 0-1 YEAR	-40.90 (1.98)	-34.28 (5.85)	5.47 (2.07)	4.99 (3.58)
YOUNGEST CH. 1-2 YEARS	-17.91 (1.46)	-23.02 (3.36)	15.20 (1.74)	16.27 (3.47)
YOUNGEST CH. 3-4 YEARS	-12.70 (1.00)	-21.08 (2.94)	9.17 (1.19)	15.62 (3.08)
YOUNGEST CH. 5-6 YEARS	-6.89 (0.93)	-16.84 (2.70)	7.33 (1.17)	15.40 (2.98)
YOUNGEST CH. 7-11 YEARS	-4.65 (0.72)	-14.31 (2.40)	4.23 (0.78)	14.25 (2.61)
YOUNGEST CH. 12-15 YEARS	-0.35 (0.73)	-2.16 (1.55)	0.94 (0.77)	4.48 (1.76)
EXP. CH. IN 0-3 MONTHS	-7.94 (2.15)	-25.26 (10.50)	-5.28 (1.92)	-11.19 (15.93)
EXP. CH. IN 4-6 MONTHS	-2.83 (1.84)	-5.99 (12.66)	-1.17 (2.41)	-14.67 (19.42)
INCOME	-0.09 (0.02)	-0.04 (0.05)	0.04 (0.03)	0.09 (0.19)
AGE/10	3.15 (1.56)	-2.34 (18.11)	0.55 (1.62)	4.08 (21.24)
AGE <sup>2</sup> /100	-0.74 (0.19)	0.22 (0.74)	0.17 (0.20)	0.42 (0.87)
INTERCEPT	9.65 (3.08)	-	2.01 (3.14)	-

Notes: Coefficient estimates have been multiplied by 100. Standard errors in parenthesis.

Table 9: OLS and GMM estimates of the parameters of the dynamic linear probability models of full-time and part-time work.

current dated values of age and the time dummies. Thus, when estimating this equation, income and fertility are assumed to be endogenous. While deeper lags of employment outcomes, income and child variables should also provide valid instruments, further moment conditions are not exploited as the Sargan statistics obtained when utilizing a large number of over-identifying restrictions tend to under reject when testing the validity of the instrument set or other relevant hypotheses (see Bowsher, 2002). The instrument set used when estimating the difference dynamic linear probability model of part-time work is identical to that used to estimate the corresponding equation for full-time work except that fertility is allowed to be predetermined rather than endogenous. Therefore the instruments obtained from the above described lags of employment outcomes and income and the instruments based on age and the time dummies are supplemented by instruments constructed from the

	FULL-TIME WORK		PART-TIME WORK	
	OLS LEVELS	DIFF. GMM	OLS LEVELS	DIFF. GMM
Sargan Test - $\chi^2(276)/\chi^2(283)$	-	282.99 $p=0.373$	-	295.44 $p=0.293$
Test for AR(1) in residuals	-3.88 $p=0.000$	-18.37 $p=0.000$	-3.46 $p=0.000$	-21.25 $p=0.000$
Test for AR(2) in residuals	-7.17 $p=0.000$	-1.07 $p=0.283$	-8.25 $p=0.000$	-0.67 $p=0.500$
Test for AR(3) in residuals	2.29 $p=0.022$	0.64 $p=0.524$	2.17 $p=0.031$	0.30 $p=0.765$
F test for joint significance	1963.53 $p=0.000$	30.15 $p=0.000$	610.77 $p=0.000$	24.38 $p=0.000$
R <sup>2</sup>	65.97%	-	53.03%	-
F test for $Y_{i,f,t-3}$ & $Y_{i,p,t-3}$	69.57 $p=0.000$	1.84 $p=0.159$	54.58 $p=0.000$	1.55 $p=0.213$
# PARAMETERS	31	26	31	26
# INSTRUMENTS	-	250	-	250
# OBSERVATIONS - $\sum_{i=1}^N T_i$	21962	17299	21962	17299
DIFFERENCE SARGAN TESTS				
Validity of most recent lag of income	-	8.39 $p=0.589$	-	6.45 $p=0.771$
Validity of most recent lag of child variables	-	67.69 $p=0.555$	-	65.60 $p=0.630$
Validity of income dated $t - 1$	-	12.55 $p=0.249$	-	16.16 $p=0.094$
Validity of child variables dated $t - 1$	-	94.80 $p=0.025$	-	-

Notes: Difference Sargan test statistics for income and child variables are distributed  $\chi^2(10)$  and  $\chi^2(70)$  respectively. Tests statistics for residual autocorrelation have standard normal distributions.

Table 10: Specification checks for the dynamic linear probability models of full-time and part-time work.

child variables dated  $t - 1$ ,  $t - 2$  and  $t - 3$ .

Sargan and difference Sargan tests reported in Table 10 support the above described choice of instrument sets. The Sargan statistics based on the first difference equations for full-time and part-time work do not reject the validity of the respective instrument sets. Also, various groups of instruments consisting of subsets of the full instrument sets are not rejected. Specifically, in the case of part-time work, instruments based on child variables dated  $t - 1$  are found to be valid instruments for the first difference equation, thus confirming that fertility is predetermined rather than endogenous with respect to part-time work. Furthermore, income dated  $t - 2$  is found to be a valid instrument for both first difference equations, while child variables dated  $t - 2$  are valid instruments for the first difference equation for full-time work. Additional instruments for the first difference equation for part-time work constructed from income dated  $t - 1$  are rejected, while child variables dated  $t - 1$  are not valid instruments for the first difference equation for full-time work. Lastly, it is noted that while additional instruments based on income dated  $t - 1$  are valid instruments for the first difference equation for full-time work these instruments are not used as when included the difference Sargan test rejects their validity.

Arellano and Bond (1991) tests for residual autocorrelation also suggest that the first

difference models of full-time and part-time work are correctly specified. As expected, tests for autocorrelation based on the residuals obtained from estimation of the first difference equations show significant negative first order autocorrelation. Importantly, there is no significant evidence of autocorrelation in higher order differences of the residuals. Thus, the first difference transformation eliminates all serial correlation in the time-varying component of unobservables. In contrast, there is significant evidence of first order and higher order autocorrelation in the residuals from the Ordinary Least Squares estimation of the two levels equations. Persistent unobservables therefore appear to be present in the untransformed equations. Given the dynamic structure of the models, ignoring these persistent unobservables will lead to biased parameter estimates. Further biases will be present if persistent unobservables are correlated with the explanatory variables.

Due to the linear structure of the models, all parameters can be interpreted as marginal effects. For example, the coefficients on  $Y_{i,f,t-1}$  and  $Y_{i,p,t-1}$  can be interpreted as the increase in the probability of full-time or part-time work caused by a move from non-employment into, respectively, full-time employment or part-time employment one year previously. These coefficients therefore represent the combined effects of search costs, habit formation and accumulation of human capital on the woman's labor supply behavior. Estimation of the levels equation describing the probability of full-time work implies that working full-time rather than being non-employed in the previous year increases a woman's probability of working full-time in the current year by 56.91(1.01) percentage points. However, estimation of the corresponding first difference equation suggests a substantially lower effect, just 35.97(2.89) percentage points. Symmetrically, results from estimation of the level equation describing the probability of part-time work imply that working part-time rather than being non-employed in the previous year increases a woman's probability of working part-time in the current year by 50.05(1.01) percentage points. In contrast, estimation of this equation in first differences suggests that the effect is only 33.91(2.47) percentage points. Thus, as for the dynamic mixed multinomial logit model, ignoring persistent unobserved heterogeneity leads to an overstatement of the effect of previous employment behavior on a woman's labor supply behavior in the current year.

The estimated effects of most individual characteristics on a woman's employment behavior are also sensitive to the modeling of persistent unobservables. Most notably, the Generalized Method of Moments estimates of the effects of fertility expectations and children aged over one year are larger in magnitude than the corresponding Ordinary Least Squares

estimates. Also, whereas the Ordinary Least Squares estimates suggest that pre-school aged children significantly increase a woman's probability of full-time work and decrease her probability of part-time work, the Generalized Method of Moments results suggest no significant effect of pre-school aged children on a woman's likelihood of working either full-time or part-time. These difference again illustrate the sensitivity of estimation results to the treatment of unobserved heterogeneity.

Examining the results from estimation of the two first difference equations reveals significant own state effects over a two year time horizon: working full-time two years previously increases a woman's probability of working full-time in the current year by 9.05(1.53) percentage points, while working part-time increases a woman's probability of being in part-time employment in the current year by 8.17(1.69) percentage points. However, for both equations, F tests for the joint significance of the woman's employment outcomes dated three years previously do not reject the null hypothesis that these additional variables are insignificant. Thus there are state dependencies in women's labor supply behavior spanning two years, but not longer. Over a one year horizon there are negative cross state effects between full-time and part-time employment. In particular, working part-time rather than being non-employed decreases a woman's probability of working full-time in the current year by 3.83(1.64) percentage points. Similarly, working full-time rather than being non-employed decreases a woman's probability of working part-time in the current year by 7.21(2.81) percentage points.

While the dynamic mixed multinomial logit models and the dynamic linear probability models rest on different approximations and are estimated using vastly different distributional assumptions, there are some similarities in the results. In terms of the implied state dependencies, Generalized Method of Moments estimation of the first difference linear probability models implies own state and cross state dependencies that are broadly comparable to those based on the results from Specification II of the dynamic mixed multinomial logit model. For example, Specification II of the dynamic mixed multinomial logit model implies that working full-time rather than being non-employed in the previous year increases a woman's probability of working full-time in the current year by 31.37(1.74) percentage points and decreases her probability of working part-time in the current year by 2.37(1.96) percentage points. The corresponding figures based on the dynamic linear probability models are 35.97(2.89) percentage points and 3.83(1.64) percentage points. These similarities suggest that the random effects assumption imposed when estimating the dynamic mixed

multinomial logit models is not overly restrictive. The similarity in the results obtained from the dynamic mixed multinomial logit model and the dynamic linear probability models diminishes as the distribution of unobservables used when estimating the dynamic mixed multinomial logit model is generalized. This finding is not surprising as the dynamic linear probability models do not allow random coefficients or autocorrelation, as is permitted in the more general specifications of the dynamic mixed multinomial logit model.

## 7 Conclusion

The high levels of persistence observed in both full-time and part-time employment has been found to be due to a combination the effect of a woman's previous employment behavior on her current labor supply problem and the effect of persistent individual characteristics. There is evidence that persistent unobserved individual characteristics contribute significantly to the observed persistence in labor supply behavior; completely ignoring persistent unobservables leads to substantial upward biases in estimated state dependencies. Moreover, different specifications of unobservables generate substantially different conclusions regarding the extent of state dependence in employment behavior. In particular, working with a specification of unobservables allowing time invariant individual specific random effects, but no further generality in the distribution of unobservables, results in significant downward biases in the estimated effect of a woman's previous employment behavior on her current choice between full-time work, part-time work and non-employment. The biases caused by imposing overly restrictive distributions of unobservables are large enough to make the choice of distribution of unobservables important when conducting policy evaluation.

One aspect of this work that has not been emphasized is the methodology used to estimate the dynamic mixed multinomial logit models. Reliability of the parameter estimates is ensured by requiring that parameters describing the intratemporal distribution of unobservables be identified solely by their effect on the intertemporal distribution of unobservables. Monte Carlo simulations show that Maximum Simulated Likelihood estimates of the parameters of dynamic mixed multinomial logit models with distributions of unobservables satisfying this requirement have desirable empirical properties. The same approach could be applied in other multinomial choice problems in labor economics, such as when modeling occupational outcomes or the type of training a worker receives, and also beyond labor economics, for example in health or education. The methodology can be applied to both



static and dynamic multinomial choice problems and is effective with only a small number of repeated observations.

Several issues raised in this paper warrant further investigation. The most substantive and interesting issues surround the implementation of the methodology. While all of the analysis conducted here has used Maximum Simulated Likelihood estimation, Method of Moments estimation or Method of Simulated Moments estimation provide potentially viable alternatives. It would be both interesting and practically useful to establish the relative merits, in terms of the associated computational burdens, accuracy and stability, of Maximum Likelihood and Method of Moments approaches to this problem. Finally, the best method of implementing simulation based estimators remains an open issue; several recent studies have illustrated that quasi random draws can improve the accuracy of Maximum Simulated Likelihood estimators (see Bhat, 2001; Train, 2001). The performance of such methods in the context of the current problem is left to future research.

## Appendix I: Monte Carlo Evidence on the Fragile Identification Problem

Monte Carlo simulations are used to illustrate the nature and severity of the fragile identification affecting the mixed multinomial logit model and to show that reliable parameter estimates are obtained if the restrictions discussed in Section 3.3 are imposed on the distribution of unobservables. To maintain consistency, attention is restricted to the three alternative model of employment dynamics described above, however similar results are obtained for static models and models with more than three alternatives.

The following specification of preferences is adopted for  $t = 3, \dots, T$

$$\begin{aligned} V_t^f(\Omega_{t-1}, Z_t) - V_t^n(\Omega_{t-1}, Z_t) &= \gamma_{f,f-1}Y_{i,f,t-1} + \gamma_{f,p-1}Y_{i,p,t-1} + \gamma_{f,f-2}Y_{i,f,t-2} + \gamma_{f,p-2}Y_{i,p,t-2} \\ &\quad + \beta_{f,0} + \beta_{f,1}x1_{i,t} + \beta_{f,2}x2_{i,t} + \eta_{i,f,t} + \xi_{i,f,t}, \end{aligned} \quad (31)$$

$$\begin{aligned} V_t^p(\Omega_{t-1}, Z_t) - V_t^n(\Omega_{t-1}, Z_t) &= \gamma_{p,f-1}Y_{i,f,t-1} + \gamma_{p,p-1}Y_{i,p,t-1} + \gamma_{p,f-2}Y_{i,f,t-2} + \gamma_{p,p-2}Y_{i,p,t-2} \\ &\quad + \beta_{p,0} + \beta_{p,1}x1_{i,t} + \beta_{p,2}x2_{i,t} + \eta_{i,p,t} + \xi_{i,p,t}. \end{aligned} \quad (32)$$

$Y_{i,j,t}$  for  $j = f, p$  are indicators of previous employment outcomes, as defined above and  $x1_{i,t}$  and  $x2_{i,t}$  are individual specific variables. In the simulations,  $x1_{i,t}$  and  $x2_{i,t}$  are constructed to be mutually independent, independent over time and individuals and to have standard normal distributions. Individuals' employment outcomes at  $t = 1$  and  $t = 2$  are determined randomly and are constructed to be independent of subsequent employment outcomes. This allows the initial conditions problem to be ignored. The unobservables  $\xi_{i,f,t}$  and  $\xi_{i,p,t}$  satisfy assumptions **A1-A4** described in Section 3.1.  $(\eta_{i,f,t}, \eta_{i,p,t})$  are assumed to be individual specific random effects. Monte Carlo simulations are conducted with two different distributions of  $(\eta_{i,f,t}, \eta_{i,p,t})$ .

In the first set of Monte Carlo simulations,  $(\eta_{i,f,t}, \eta_{i,p,t})$  have the following distribution

$$\eta_{i,f,t} = \nu_{i,f} \quad \text{for } t = 3, \dots, T, \quad (33)$$

$$\eta_{i,p,t} = \nu_{i,p} \quad \text{for } t = 3, \dots, T, \quad (34)$$

where  $(\nu_{i,f}, \nu_{i,p})' \sim N(0, \Sigma)$ . This specification of unobservables satisfies the requirement discussed in Section 3.3 and therefore should yield well behaved parameter estimates provided  $T \geq 4$ .

In the second set of simulations,  $(\eta_{i,f,t}, \eta_{i,p,t})$  have the following distribution

$$\eta_{i,f,t} = \nu_{i,f} + \sum_{t=3}^T \pi_{i,f,t} I_t \quad \text{for } t = 3, \dots, T, \quad (35)$$

$$\eta_{i,p,t} = \nu_{i,p} + \sum_{t=3}^T \pi_{i,p,t} I_t \quad \text{for } t = 3, \dots, T. \quad (36)$$

Again,  $(\nu_{i,f}, \nu_{i,p})' \sim N(0, \Sigma)$ .  $I_t$  for  $t = 3, \dots, T$  are time dummies and  $(\pi_{i,f,t}, \pi_{i,p,t})$  for  $t = 3, \dots, T$  are random coefficients that are independent over time and individuals with  $(\pi_{i,f,t}, \pi_{i,p,t})' \sim N(0, \Xi^t)$  for  $t = 3, \dots, T$ . When estimating this model, normalizations are imposed on  $\Xi_{1,1}^t$  for  $t = 3, \dots, T$ . Without such normalizations, scale identification relies on the slight difference in the shapes of the logistic and normal distributions (see Ben-Akiva *et al.*, 2001). This specification of unobservables does not satisfy the above described requirement as the parameters appearing in  $\Xi^t$  for  $t = 3, \dots, T$  affect the distribution of unobservables at time  $t$  but do not affect the intertemporal distribution of unobservables. Thus, while all parameters in this specification are identified, the model is likely to suffer from the fragile identification problem.

The two sets of Monte Carlo simulations were conducted with sample sizes of 3000 individuals and  $T = 4$ . In each Monte Carlo experiment, 200 data sets were generated and Maximum Simulated Likelihood estimates obtained for each data set. Table 11 reports the results. In the first Monte Carlo simulation, average parameter estimates correspond closely to their true values and average standard errors are almost identical to the standard deviation of the parameter estimates. Convergence was obtained for all of the 200 Monte Carlo replications, and took an average of 4.18 iterations starting from the true parameter values. Thus, for this specification of unobservables, parameter estimates and standard errors are reliable and there is no evidence of numerical instability.

In contrast, the Monte Carlo results for the second specification of unobservables reveal major problems. In many cases, the average coefficients on the explanatory variables differ substantially from their true values, and average standard errors bear little resemblance to the standard deviation of the parameter estimates. The estimates of the parameters of the covariance matrices reveal even greater problems: in many cases average variances are many times greater than their true values and average standard errors are huge. Furthermore, in around 10% of the Monte Carlo replications, convergence was not obtained within the first 200 iterations.

PARAMETER	TRUTH	MONTE CARLO SIMULATION I			MONTE CARLO SIMULATION II		
		E(parameter)	E( $\sigma$ )	$\sigma$ (parameter)	E(parameter)	E( $\sigma$ )	$\sigma$ (parameter)
$\gamma_{f,f-2}$	1	0.99	0.14	0.14	0.96	0.17	0.15
$\gamma_{f,p-2}$	0.5	0.48	0.14	0.13	0.49	0.22	0.25
$\gamma_{f,f-1}$	2	2.02	0.15	0.15	2.12	0.20	0.22
$\gamma_{f,p-1}$	1	1.00	0.14	0.14	1.11	0.32	0.45
$\beta_{f,0}$	-1	-1.00	0.17	0.18	-1.03	0.51	0.68
$\beta_{f,1}$	-0.8	-0.80	0.09	0.09	-0.78	0.23	0.31
$\beta_{f,2}$	0.5	0.50	0.07	0.07	0.48	0.14	0.18
$\gamma_{p,f-2}$	0.5	0.51	0.12	0.11	0.34	0.63	0.56
$\gamma_{p,p-2}$	1	0.99	0.13	0.11	1.71	2.19	1.76
$\gamma_{p,f-1}$	1	1.02	0.14	0.12	0.91	0.68	0.51
$\gamma_{p,p-1}$	2	2.01	0.12	0.13	3.60	4.63	3.82
$\beta_{p,0}$	0.5	0.50	0.13	0.13	0.41	0.49	0.43
$\beta_{p,1}$	1	1.01	0.08	0.08	2.59	4.63	3.85
$\beta_{p,2}$	-0.5	-0.51	0.06	0.06	-1.39	2.58	2.18
$\Sigma_{1,1}$	1	1.01	0.40	0.39	0.97	0.57	0.56
$\Sigma_{2,1}$	0.5	0.51	0.27	0.27	0.49	1.15	0.82
$\Sigma_{2,2}$	1	1.06	0.33	0.33	11.81	63.40	40.72
$\Xi_{1,1}^3$	4 [Fixed]	-	-	-	4	-	-
$\Xi_{2,1}^3$	1	-	-	-	-0.83	7.46	7.70
$\Xi_{2,2}^3$	2	-	-	-	59.19	314.23	171.98
$\Xi_{1,1}^4$	4 [Fixed]	-	-	-	4	-	-
$\Xi_{2,1}^4$	1	-	-	-	-0.40	6.46	6.10
$\Xi_{2,2}^4$	2	-	-	-	57.53	313.81	178.98
AVERAGE ITERATIONS			4.18			38.41	
MAXIMUM ITERATIONS			10			200	

Notes: E(parameter) is the mean parameter estimate, E( $\sigma$ ) is the mean estimated standard error and  $\sigma$ (parameter) is the standard deviation of the parameter estimates over the 200 Monte Carlo replications. Maximum Simulated Likelihood estimation used 5000 antithetic draws. The number of iterations is limited to 200.

Table 11: Monte Carlo simulations illustrating the fragile identification problem and the effectiveness of the solution proposed in Section 3.3.

## Appendix II: The Performance of Estimators of the Parameters in Dynamic Mixed Multinomial Logit Models

Two further Monte Carlo simulations are conducted in order to establish the empirical properties of the Maximum Simulated Likelihood estimator in the context of dynamic mixed multinomial logit models with unobservables as in Specifications V and VI, described above in Section 5. Payoffs are as described by Equations (31) and (32). For each of the specifications of unobservables, 100 data sets were generated each with the same sample size, attrition pattern and distribution of the initial conditions as observed in the BHPS sample. The likelihood functions were simulated using 5000 antithetic draws.

The results of Monte Carlo the simulations are summarized in Tables 12 and 13. There is a close correspondence between the average estimates of the deterministic components of coefficients and the true values, and the average standard errors are close to the standard

deviation of the parameter estimates. There is, however, evidence of biases in some of the parameters appearing in the distribution of unobservables. In particular, some of the variances of the random coefficients appear to be biased downwards. These biases are due to finite sample bias and simulation noise. Tables 14 shows the average dynamic responses to policy interventions evaluated at the estimated parameter values and at the true parameter values. On average, over the 100 Monte Carlo replications, the dynamic responses evaluated at the estimated parameters are very close to the dynamic responses evaluated at the true parameter values. It appears therefore that moderate biases in parameters translate into very small biases in dynamic responses.

VARIABLE	TRUTH		SPEC. V		SPEC VI	
	$f$	$p$	$f$	$p$	$f$	$p$
$Y_{i,f,t-2}$	1.00	0.50	0.99 (0.13)[0.12]	0.50 (0.15)[0.13]	0.99 (0.12)[0.12]	0.48 (0.12)[0.12]
$Y_{i,p,t-2}$	0.50	1.00	0.53 (0.12)[0.11]	1.03 (0.12)[0.11]	0.50 (0.11)[0.12]	0.97 (0.11)[0.11]
$Y_{i,f,t-1}$	2.00	1.00	1.97 (0.14)[0.13]	0.96 (0.14)[0.11]	2.01 (0.14)[0.18]	1.01 (0.13)[0.15]
$Y_{i,p,t-1}$	1.00	2.00	0.95 (0.12)[0.12]	1.95 (0.14)[0.13]	1.01 (0.13)[0.13]	1.98 (0.12)[0.10]
$x1_{i,t}$	-0.80	1.00	-0.80 (0.05)[0.05]	1.00 (0.05)[0.05]	-0.74 (0.06)[0.05]	0.99 (0.05)[0.05]
$x2_{i,t}$	0.50	-0.50	0.50 (0.04)[0.04]	-0.50 (0.05)[0.04]	0.47 (0.04)[0.05]	-0.50 (0.04)[0.05]

Notes: Average standard errors are given in round brackets and the standard deviation of the parameter estimates is given in square brackets. Estimates of the parameters on the initial conditions are omitted. Results are based on 100 Monte Carlo replications.

Table 12: Results of Monte Carlo simulations I: Average estimates of the deterministic components of coefficients.

	MONTE CARLO FOR SPEC. V		MONTE CARLO FOR SPEC. VI	
	TRUTH	RESULTS	TRUTH	RESULTS
$\Sigma_{Intercept}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.91 & \cdot \\ (0.28)[0.47] & \\ 0.46 & 0.99 \\ (0.21)[0.31] & (0.26)[0.31] \end{pmatrix}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.81 & \cdot \\ (0.32)[0.48] & \\ 0.41 & 0.70 \\ (0.23)[0.34] & (0.29)[0.47] \end{pmatrix}$
$\rho_f$			0.7	0.72 (0.09)[0.15]
$\rho_p$			0.8	0.85 (0.05)[0.06]
$\Sigma_{\zeta 1,1}$			2	1.51 (0.48)[0.61]
$\Sigma_{\zeta 2,1}$			0.7	0.72 (0.10)[0.13]
$\Sigma_{\zeta 2,2}$			2	1.97 (0.46)[0.62]
$\Sigma_{Y_{i,f,t-2}}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 1.01 & \cdot \\ (0.36)[0.43] & \\ 0.49 & 0.96 \\ (0.30)[0.36] & (0.37)[0.44] \end{pmatrix}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.77 & \cdot \\ (0.34)[0.37] & \\ 0.41 & 0.95 \\ (0.29)[0.34] & (0.37)[0.43] \end{pmatrix}$
$\Sigma_{Y_{i,p,t-2}}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 1.04 & \cdot \\ (0.35)[0.40] & \\ 0.56 & 1.05 \\ (0.26)[0.30] & (0.28)[0.27] \end{pmatrix}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.83 & \cdot \\ (0.35)[0.40] & \\ 0.43 & 0.94 \\ (0.26)[0.28] & (0.28)[0.30] \end{pmatrix}$
$\Sigma_{Y_{i,f,t-1}}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.82 & \cdot \\ (0.36)[0.35] & \\ 0.39 & 0.89 \\ (0.32)[0.29] & (0.39)[0.40] \end{pmatrix}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.93 & \cdot \\ (0.38)[0.45] & \\ 0.45 & 0.96 \\ (0.32)[0.46] & (0.39)[0.64] \end{pmatrix}$
$\Sigma_{Y_{i,p,t-1}}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.91 & \cdot \\ (0.35)[0.39] & \\ 0.39 & 0.88 \\ (0.25)[0.29] & (0.27)[0.32] \end{pmatrix}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.86 & \cdot \\ (0.36)[0.37] & \\ 0.44 & 0.92 \\ (0.27)[0.27] & (0.30)[0.29] \end{pmatrix}$
$\Sigma_{x_{i,1,t}}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.99 & \cdot \\ (0.13)[0.15] & \\ 0.50 & 1.01 \\ (0.09)[0.09] & (0.12)[0.12] \end{pmatrix}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.93 & \cdot \\ (0.14)[0.15] & \\ 0.48 & 0.95 \\ (0.10)[0.11] & (0.12)[0.14] \end{pmatrix}$
$\Sigma_{x_{i,2,t}}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.96 & \cdot \\ (0.13)[0.14] & \\ 0.47 & 0.96 \\ (0.09)[0.11] & (0.11)[0.13] \end{pmatrix}$	$\begin{pmatrix} 1 & \cdot \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 0.91 & \cdot \\ (0.13)[0.13] & \\ 0.46 & 0.94 \\ (0.09)[0.09] & (0.12)[0.12] \end{pmatrix}$

Notes: Average standard errors are given in round brackets and the standard deviation of the parameter estimates is given in square brackets. Results are based on 100 Monte Carlo replications.

Table 13: Results of Monte Carlo simulations II: Average estimates of parameters in the distribution of unobservables.

EMPLOYMENT STATE	YEARS SINCE INTERVENTION										
	1	2	3	4	5	6	7	8	9	10	11
SPECIFICATION V											
POLICY INCENTIVIZING NON-EMPLOYED WOMEN TO MOVE INTO FULL-TIME WORK											
<i>f</i>	14.94 (14.96)	10.80 (10.83)	3.95 (3.99)	3.33 (3.42)	1.77 (1.76)	1.39 (1.44)	0.85 (0.87)	0.70 (0.75)	0.47 (0.48)	0.37 (0.39)	0.30 (0.30)
<i>p</i>	-3.47 (-3.34)	-4.60 (-4.62)	-2.17 (-2.20)	-2.29 (-2.33)	-1.33 (-1.32)	-1.13 (-1.16)	-0.72 (-0.71)	-0.61 (-0.64)	-0.42 (-0.41)	-0.34 (-0.34)	-0.28 (-0.27)
<i>n</i>	-11.48 (-11.62)	-6.20 (-6.21)	-1.78 (-1.78)	-1.04 (-1.09)	-0.44 (-0.44)	-0.25 (-0.28)	-0.13 (-0.16)	-0.09 (-0.12)	-0.05 (-0.06)	-0.04 (-0.05)	-0.02 (-0.03)
POLICY INCENTIVIZING NON-EMPLOYED WOMEN TO MOVE INTO PART-TIME WORK											
<i>f</i>	-3.61 (-3.54)	-3.32 (-3.36)	-1.44 (-1.42)	-1.60 (-1.62)	-0.89 (-0.89)	-0.73 (-0.73)	-0.46 (-0.44)	-0.39 (-0.38)	-0.27 (-0.30)	-0.23 (-0.24)	-0.16 (-0.18)
<i>p</i>	16.41 (16.49)	10.79 (10.79)	3.68 (3.67)	3.09 (3.10)	1.53 (1.52)	1.17 (1.17)	0.68 (0.68)	0.55 (0.55)	0.36 (0.37)	0.30 (0.31)	0.19 (0.23)
<i>n</i>	-12.80 (-12.95)	-7.47 (-7.44)	-2.24 (-2.25)	-1.49 (-1.47)	-0.64 (-0.63)	-0.43 (-0.43)	-0.21 (-0.24)	-0.16 (-0.17)	-0.09 (-0.07)	-0.07 (-0.08)	-0.03 (-0.05)
SPECIFICATION VI											
POLICY INCENTIVIZING NON-EMPLOYED WOMEN TO MOVE INTO FULL-TIME WORK											
<i>f</i>	15.13 (14.19)	10.92 (10.18)	3.95 (3.50)	3.14 (2.95)	1.60 (1.44)	1.23 (1.16)	0.79 (0.68)	0.59 (0.53)	0.41 (0.36)	0.34 (0.30)	0.24 (0.22)
<i>p</i>	-2.94 (-2.61)	-3.89 (-3.51)	-1.73 (-1.50)	-1.95 (-1.78)	-1.08 (-0.93)	-0.93 (-0.86)	-0.61 (-0.53)	-0.50 (-0.42)	-0.35 (-0.30)	-0.29 (-0.26)	-0.22 (-0.19)
<i>n</i>	-12.19 (-11.57)	-7.03 (-6.67)	-2.22 (-2.01)	-1.19 (-1.17)	-0.52 (-0.51)	-0.30 (-0.30)	-0.18 (-0.15)	-0.09 (-0.11)	-0.06 (-0.06)	-0.05 (-0.05)	-0.03 (-0.02)
POLICY INCENTIVIZING NON-EMPLOYED WOMEN TO MOVE INTO PART-TIME WORK											
<i>f</i>	-2.97 (-3.21)	-2.70 (-2.88)	-1.13 (-1.20)	-1.37 (-1.38)	-0.70 (-0.73)	-0.61 (-0.62)	-0.37 (-0.42)	-0.31 (-0.34)	-0.22 (-0.25)	-0.16 (-0.16)	-0.12 (-0.12)
<i>p</i>	16.22 (16.11)	10.76 (10.77)	3.72 (3.71)	3.06 (3.02)	1.44 (1.47)	1.11 (1.09)	0.65 (0.65)	0.49 (0.49)	0.33 (0.34)	0.25 (0.24)	0.17 (0.17)
<i>n</i>	-13.25 (-12.90)	-8.06 (-7.90)	-2.59 (-2.51)	-1.69 (-1.64)	-0.74 (-0.74)	-0.50 (-0.47)	-0.28 (-0.23)	-0.18 (-0.14)	-0.11 (-0.10)	-0.09 (-0.07)	-0.05 (-0.05)

Notes: For each Monte Carlo replication, average policy effects for the sample individuals were obtained using the estimated parameters and the true parameters. The figures in this table are averages of these quantities over the 100 Monte Carlo replications. All figures are percentage point changes for women affected by the policy.

Table 14: Results of Monte Carlo simulations III: Average policy responses evaluated at the estimated parameters. Average true policy responses are shown in parenthesis.

## Appendix III: Goodness of Fit of the Dynamic Mixed Multinomial Logit Model

EMPLOYMENT STATE	YEAR										
STATE	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
	OBSERVED										
<i>f</i>	42.46	43.55	44.38	45.43	45.24	45.82	46.00	45.68	46.16	45.27	44.27
<i>p</i>	30.19	29.60	29.23	29.84	29.72	28.70	29.84	29.16	28.44	29.88	31.00
<i>n</i>	27.34	26.85	26.39	24.74	25.03	25.48	24.16	25.16	25.40	24.85	24.73
	SPECIFICATION I										
<i>f</i>	42.47	43.46	44.25	45.16	45.15	45.65	45.73	45.55	45.94	45.05	44.08
<i>p</i>	30.21	29.40	28.98	29.43	29.33	28.54	29.86	28.97	28.29	29.59	30.79
<i>n</i>	27.33	27.14	26.77	25.41	25.52	25.81	24.41	25.49	25.77	25.36	25.14
	SPECIFICATION II										
<i>f</i>	42.56	43.74	44.60	45.53	45.36	45.83	45.86	45.74	45.89	44.99	44.09
<i>p</i>	30.05	29.39	28.75	29.39	29.29	28.46	29.60	28.82	28.24	29.51	30.59
<i>n</i>	27.39	26.87	26.64	25.08	25.36	25.70	24.54	25.44	25.87	25.50	25.32
	SPECIFICATION III										
<i>f</i>	42.53	43.68	44.50	45.38	45.34	45.79	45.76	45.64	45.92	44.95	44.08
<i>p</i>	30.07	29.43	28.80	29.41	29.13	28.29	29.54	28.73	28.05	29.41	30.42
<i>n</i>	27.40	26.89	26.70	25.21	25.53	25.92	24.71	25.63	26.03	25.64	25.49
	SPECIFICATION IV										
<i>f</i>	42.60	43.69	44.62	45.66	45.56	45.98	45.95	45.78	46.11	45.17	44.37
<i>p</i>	29.87	29.35	28.60	29.04	28.95	28.31	29.58	28.78	27.96	29.34	30.26
<i>n</i>	27.53	26.96	26.78	25.31	25.49	25.71	24.47	25.44	25.93	25.48	25.36
	SPECIFICATION V										
<i>f</i>	42.50	43.76	44.60	45.56	45.46	46.17	45.94	45.74	45.92	45.02	43.96
<i>p</i>	30.05	29.40	28.71	29.10	28.90	28.08	29.30	28.55	28.04	29.30	30.45
<i>n</i>	27.45	26.84	26.69	25.33	25.64	25.75	24.75	25.70	26.04	25.69	25.59
	SPECIFICATION VI										
<i>f</i>	43.27	42.00	42.61	43.77	44.41	45.26	45.14	45.69	45.67	45.63	45.81
<i>p</i>	29.71	29.37	30.03	29.28	28.64	29.11	29.04	28.48	29.63	28.81	28.24
<i>n</i>	27.02	28.63	27.36	26.94	26.95	25.64	25.83	25.82	24.70	25.56	25.95

Notes: Observed percentages are computed from the sample observations. The remaining figures are simulation based predictions using the results from the six specifications of the dynamic mixed multinomial logit model.

Table 15: Goodness of fit I: Predicted percentage of sampled women choosing each employment state in each year 1993-2003.



EMPLOY. STATE	TIME												TOTAL ABS. DIFF.
	$t-1$	$t-2$	$t-3$	$t-4$	$t-5$	$t-6$	$t-7$	$t-8$	$t-9$	$t-10$	$t-11$	$t-12$	
OBSERVED													
<i>f</i>	86.97	82.27	78.40	74.53	71.86	69.58	67.05	64.60	61.97	60.83	58.97	56.42	
<i>p</i>	78.40	72.39	68.23	65.35	63.16	60.28	58.57	55.94	54.05	52.38	52.09	51.04	0.00
<i>n</i>	78.88	72.76	68.07	64.02	60.88	57.53	54.71	51.15	48.08	44.25	43.05	41.63	
SPECIFICATION I													
<i>f</i>	86.75	81.89	76.25	71.81	68.10	64.84	62.19	60.02	58.45	57.27	56.78	56.04	
<i>p</i>	78.22	72.28	65.44	60.83	57.44	54.61	52.58	51.27	49.97	49.29	48.75	48.22	107.95
<i>n</i>	78.74	72.80	65.36	60.09	55.52	52.04	49.25	46.92	45.49	44.26	43.96	43.21	
SPECIFICATION II													
<i>f</i>	86.80	82.03	77.37	73.93	71.36	68.91	66.85	65.23	63.93	62.83	61.93	61.27	
<i>p</i>	78.01	72.01	66.41	62.88	60.61	58.48	56.91	55.78	54.54	53.86	52.86	52.33	74.57
<i>n</i>	78.86	73.09	67.74	64.28	61.50	59.39	57.53	55.76	54.53	53.00	52.09	50.98	
SPECIFICATION III													
<i>f</i>	86.81	82.01	77.33	73.85	71.26	68.86	66.83	65.08	63.86	62.71	61.57	60.70	
<i>p</i>	78.10	72.20	66.61	62.98	60.51	58.35	56.85	55.52	54.64	53.90	53.20	52.21	78.80
<i>n</i>	78.86	73.06	67.77	64.50	61.59	59.63	57.95	56.46	55.16	53.63	52.83	52.04	
SPECIFICATION IV													
<i>f</i>	86.79	82.09	77.49	74.19	71.67	69.24	67.24	65.48	64.08	62.93	62.04	61.40	
<i>p</i>	77.91	71.88	66.36	62.89	60.67	58.55	57.06	55.84	54.59	53.73	52.76	52.00	72.50
<i>n</i>	78.81	73.05	67.59	64.14	61.33	59.22	57.46	55.77	54.59	52.83	51.90	50.53	
SPECIFICATION V													
<i>f</i>	86.76	81.89	77.09	73.61	70.95	68.51	66.48	64.76	63.44	62.29	61.47	60.93	
<i>p</i>	78.19	72.07	66.48	63.03	60.78	58.62	56.93	55.58	54.06	53.12	51.80	51.00	75.91
<i>n</i>	79.02	73.14	67.89	64.76	62.07	60.01	58.10	56.36	55.17	53.67	52.64	51.63	
SPECIFICATION VI													
<i>f</i>	86.88	82.03	77.21	73.51	70.64	68.03	65.92	64.04	62.73	61.52	60.66	60.49	
<i>p</i>	78.19	72.09	66.45	62.88	60.48	58.33	56.69	55.34	54.04	53.18	52.16	51.45	62.58
<i>n</i>	79.14	73.32	68.07	64.82	61.95	59.79	57.86	56.32	55.23	53.67	52.88	51.58	

Notes: Total absolute difference is the sum of the absolute differences between the observed and predicted percentages. Also see the notes for Table 15.

Table 16: Goodness of fit II: Observed and predicted percentage of women working full-time or part-time or in non-employment in year  $t - j$  who are in the same employment state in year  $t$ .

## References

- ADAMS, J., CLARK, M., EZROW, L. AND GLASGOW, G. (1999), Are Niche Parties Fundamentally Different from Mainstream Parties? The Causes and the Electoral Consequences of Western European Parties' Policy Shifts, 1976-1998, *American Journal of Political Science*, **vol. 50**(3): pp. 513–529.
- AKAIKE, H. (1973), Information Theory as an Extension of the Maximum Likelihood Principle, in B. Petrov and F. Csaki (Editors), *Second International Symposium on Information Theory*, pp. 267–281, Akademiai Kiado, Budapest.
- ALTUG, S. AND MILLER, R. (1998), The Effect of Work Experience on Female Wages and Labour Supply, *Review of Economic Studies*, **vol. 65**(1): pp. 45–85.
- ANDERSON, T. W. AND HSIAO, C. (1982), Formulation and Estimation of Dynamic Models using Panel Data, *Journal of Econometrics*, **vol. 18**(1): pp. 47–82.
- ARELLANO, M. AND BOND, S. (1991), Some Tests of Specification for Panel Data, *Review of Economic Studies*, **vol. 58**(2): pp. 277–297.
- BEN-AKIVA, M., BOLDUC, D. AND BRADLEY, M. (1993), Estimation of Travel Choice Models with Randomly Distributed Values of Time, Papers 9303, Laval - Recherche en Energie.
- BEN-AKIVA, M., BOLDUC, D. AND WALKER, J. (2001), Specification, Identification and Estimation of the Logit Kernel (or Continuous Mixed Logit Model), *Department of Civil Engineering Manuscript, MIT*.
- BHAT, C. R. (1998), Accommodating Variations in Responsiveness to Level-of-service Measures in Travel Mode Choice Modeling, *Transportation Research Part A: Policy and Practice*, **vol. 32**(7): pp. 495–507.
- BHAT, C. R. (2001), Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model, *Transportation Research B: Methodological*, **vol. 35**(7): pp. 677–693.
- BINGLEY, P. AND WALKER, I. (1997), The Labour Supply, Unemployment and Participation of Lone Mothers in In-Work Transfer Programmes, *Economic Journal*, **vol. 107**(444): pp. 1375–1390.
- BLUNDELL, R. AND BOND, S. (1998), Initial Conditions and Moment Conditions in Dynamic Panel Data Models, *Journal of Econometrics*, **vol. 87**(1): pp. 115–143.
- BOLDUC, D. (1992), Generalized Autoregressive Errors in the Multinomial Probit Model, *Transportation Research B: Methodological*, **vol. 26**(2): pp. 155–170.
- BOOTH, A., JENKINS, S. AND SERRANO, C. (1999), New Men and New Women? A

- Comparison of Paid Work Propensities from a Panel Data Perspective, *Oxford Bulletin of Economics and Statistics*, **vol. 61**(2): pp. 167–197.
- BOUND, J., SCHOENBAUM, M., STINEBRICKNER, T. R. AND WAIDMANN, T. (1999), The Dynamic Effects of Health on the Labor Force Transitions of Older Workers, *Labour Economics*, **vol. 6**(2): pp. 179–202.
- BOVER, O. (1991), Relaxing Intertemporal Separability: A Rational Habits Model of Labor Supply Estimated from Panel Data, *Journal of Labor Economics*, **vol. 9**(1): pp. 85–100.
- BOWSER, C. G. (2002), On Testing Overidentifying Restrictions in Dynamic Panel Data Models, *Economics Letters*, **vol. 77**(2): pp. 211–220.
- BRENDAN, J., DALE, A. AND JOSHI, H. (1997), Part-time Work Among British Women, in Bloosfeld and Hakin (Editors), *Between Equalization and Maginalization*, pp. 210–246, Oxford University Press.
- BROWNSTONE, D. AND TRAIN, K. (1998), Forecasting New Product Penetration with Flexible Substitution Patterns, *Journal of Econometrics*, **vol. 89**(1-2): pp. 109–129.
- BUNCH, D. S. (1991), Estimability in the Multinomial Probit Model, *Transportation Research B: Methodological*, **vol. 25**(1): pp. 1–12.
- BUNCH, D. S. AND KITAMURA, R. (1991), Multinomial Probit Model Estimation Revisited: Testing Estimable Model Specifications, Maximum Likelihood Algorithms, and Probit Integral Approximations for Trinomial Models of Household Car Ownership, University of California Transportation Centre Working Paper 70.
- CHINTAGUNTA, P. K. (1992), Estimating a Multinomial Probit Model of Brand Choice Using the Method of Simulated Moments, *Marketing Science*, **vol. 11**(4): pp. 386–407.
- DANCER, D. M. AND FIEBIG, D. G. (2004), Modelling Students at Risk, *Australian Economic Papers*, **vol. 43**(2): pp. 158–173.
- DANSIE, B. R. (1985), Parameter Estimability in the Multinomial Probit Model, *Transportation Research B: Methodological*, **vol. 19**(6): pp. 526–528.
- DUNCAN, A. AND GILES, C. (1996), Labour Supply Incentives and Recent Family Credit Reforms, *Economic Journal*, **vol. 106**(434): pp. 142–155.
- ECKSTEIN, Z. AND WOLPIN, K. (1989), The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey, *The Journal of Human Resources*, **vol. 24**(4): pp. 562–598.
- FITZGERALD, J., GOTTSCHALK, P. AND MOFFITT, R. (1998), An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics, University of Wisconsin Institute for Research on Poverty Discussion Papers, 1156-1198.
- FORMAN, C. (2005), The Corporate Digital Divide: Determinants of Internet Adoption,

- Management Science*, **vol. 51**(4): pp. 641–654.
- FRANCESCONI, M. (2002), A Joint Dynamic Model of Fertility and Work of Married Women, *Journal of Labor Economics*, **vol. 20**(2): pp. 336–380.
- GEWEKE, J. F. (1996), Monte Carlo Simulation and Numerical Integration, in K. D. Amman, H. and J. Rust (Editors), *Handbook of Computational Economics*, vol. 13, pp. 731–800, Amsterdam: North Holland.
- GEWEKE, J. F., KEANE, M. P. AND RUNKLE, D. E. (1997), Statistical Inference in the Multinomial Multiperiod Probit Model, *Journal of Econometrics*, **vol. 80**(1): pp. 125–165.
- GIANNELLI, G. C. AND MONFARDINI, C. (2003), Joint Decisions on Household Membership and Human Capital Accumulation of Youths. The Role of Expected Earnings and Local Markets, *Journal of Population Economics*, **vol. 16**(2): pp. 265–285.
- HAJIVASSILIOU, V. (1999), Some Practical Issues in Maximum Simulated Maximum Likelihood, in R. Mariano, T. Schuermann and M. Weeks (Editors), *Simulation-Based Inference in Econometrics: Methods and Applications*, chap. 3, pp. 71–99, Cambridge University Press.
- HAJIVASSILIOU, V. AND RUDD, P. A. (1994), Classical Estimation Methods for LDV Models Using Simulation, in C. Engle and D. McFadden (Editors), *Handbook of Econometrics*, pp. 2383–2441, Amsterdam: North Holland.
- HANSEN, L. (1982), Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, **vol. 50**(4): pp. 1029–1054.
- HARRIS, K. M. AND KEANE, M. P. (1998), A Model of Health Plan Choice: Inferring Preferences and Perceptions from a Combination of Revealed Preference and Attitudinal Data, *Journal of Econometrics*, **vol. 89**(1-2): pp. 131–157.
- HAUSMAN, J. AND WISE, D. (1979), Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment, *Econometrica*, **vol. 47**(2): pp. 455–473.
- HECKMAN, J. J. (1981), The Incidental Parameters Problem and the Problem of Initial Condition in Estimating a Discrete Time-Discrete Data Stochastic Process, in C. Manski and D. McFadden (Editors), *Structural Analysis of Discrete Data and Econometric Applications*, chap. 4, pp. 179–197, Cambridge: The MIT Press.
- HECKMAN, J. J. AND BORJAS, G. (1980), Does Unemployment Cause Future Unemployment? Definitions, Questions and Answers from a Continuous Time Model of Heterogeneity and State Dependence, *Economica*, **vol. 47**(127): pp. 247–283.
- HECKMAN, J. J. AND SEDLACEK, G. (1985), Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-selection in the Labor Market, *Journal of Political*

- Economy*, **vol. 93**(6): pp. 1077–1125.
- HECKMAN, J. J. AND WILLIS, R. (1977), A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Women, *Journal of Political Economy*, **vol. 85**(1): pp. 27–58.
- HOLTZ-EAKIN, D., NEWEY, W. AND ROSEN, H. (1988), Estimating Vector Autoregressions with Panel Data, *Econometrica*, **vol. 56**(6): pp. 1371–1395.
- HYSLOP, D. (1999), State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women, *Econometrica*, **vol. 67**(6): pp. 1255–1294.
- IMAI, S. AND KEANE, M. P. (2004), Intertemporal Labor Supply and Human Capital Accumulation, *International Economic Review*, **vol. 45**(2): pp. 601–641.
- KEANE, M. P. (1992), A Note on Identification in the Multinomial Probit Model, *Journal of Business and Economic Statistics*, **vol. 10**(2): pp. 193–200.
- KEANE, M. P. (1997), Modeling Heterogeneity and State Dependence in Consumer Choice Behavior, *Journal of Business & Economic Statistics*, **vol. 15**(3): pp. 310–27.
- KEANE, M. P. AND MOFFITT, R. (1998), A Structural Model of Multiple Welfare Program Participation and Labor Supply, *International Economic Review*, **vol. 39**(3): pp. 553–589.
- KEANE, M. P. AND WOLPIN, K. I. (1997), The Career Decisions of Young Men, *Journal of Political Economy*, **vol. 105**(3): pp. 473–522.
- KEANE, M. P. AND WOLPIN, K. I. (2001), The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment, *International Economic Review*, **vol. 42**(4): pp. 1051–1103.
- KUBIN, I. AND PRINZ, A. (2002), Labour Supply with Habit Formation, *Economics Letters*, **vol. 75**(1): pp. 75–79.
- LACY, D. AND BURDEN, B. C. (1999), The Vote Stealing and Turnout Effects of Ross Perot in the 1992 U.S. Presidential Election, *American Journal of Political Science*, **vol. 43**(1): pp. 233–255.
- LAYARD, R. AND BEAN, C. (1989), Why Does Unemployment Persist?, *Scandinavian Journal of Economics*, **vol. 91**(2): pp. 371–396.
- MANNING, A. AND PETRONGOLO, B. (2005), The Part-time Pay Penalty, *CEP Discussion Paper*.
- MUELLBAUER, J. (1988), Habits, Rationality and Myopia in the Life Cycle Model, *Annales d'Economie et de Statistique*, **vol. 9**: pp. 47–70.
- O'REILLY, J. AND BOTHFELD, S. (2002), What Happens After Working Part-time? Integration, Maintenance or Exclusionary Transitions in Britain and Western Germany,

- Cambridge Journal of Economics*, **vol. 26**(4): pp. 409–439.
- PROWSE, V. (2005), State Dependence in a Multi-State Model of Employment Dynamics, IZA Discussion Papers 1623, Institute for the Study of Labor (IZA).
- QUINN, K. M. AND MARTIN, A. D. (1999), Voter Choice in Multi-party Democracies: A Test of Competing Theories and Models, *American Journal of Political Science*, **vol. 43**(4): pp. 1231–1247.
- RUST, J. (1996), Numerical Dynamic Programming in Economics, in H. M. Amman, D. A. Kendrick and J. Rust (Editors), *Handbook of Computational Economics*, vol. 1, chap. 14, pp. 619–729, Amsterdam: North Holland.
- SCHWARTZ, G. (1978), Estimating the Dimension of a Model, *Annals of Statistics*, **vol. 6**: pp. 461–464.
- SHAW, K. (1989), Life-Cycle Labor Supply with Human Capital Accumulation, *International Economic Review*, **vol. 30**(2): pp. 431–456.
- STURM, R. AND WELLS, K. (1998), Physician Knowledge, Financial Incentives and Treatment Decisions for Depression, *Journal of Mental Health Policy and Economics*, **vol. 1**(2): pp. 89–100.
- TRAIN, K. (2001), Halton Sequences for Mixed Logit, *Working paper, Department of Economics, University of California, Berkeley*.
- VAN SOEST, A. (1995), Structural Models of Family Labor Supply: A Discrete Choice Approach, *Journal of Human Resources*, **vol. 30**(1): pp. 63–88.
- VAN SOEST, A., DAS, M. AND GONG, X. (2002), A Structural Labour Supply Model with Flexible Preferences, *Journal of Econometrics*, **vol. 107**(1): pp. 345–374.
- WOITTEZ, I. AND KAPTEYN, A. (1998), Social Interactions and Habit Formation in a Model of Female Labour Supply, *Journal of Public Economics*, **vol. 70**(2): pp. 185–205.
- WOLPIN, K. I. (1992), The Determinants of Black-White Differences in Early Employment Careers: Search, Layoffs, Quits, and Endogenous Wage Growth, *The Journal of Political Economy*, **vol. 100**(3): pp. 535–560.
- WOOLDRIDGE, J. M. (2005), Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity, *Journal of Applied Econometrics*, **vol. 20**(1): pp. 39–54.