Decision Structures and Discrete Choice: An Application to Labour Supply and Fertility

M. Laura Di Tommaso
Dept. Applied Economics
University of Cambridge
Austin Robinson Building
Cambridge, CB3 9DE

M. Weeks
Faculty of Economics
and Politics
University of Cambridge
Austin Robinson Building
Cambridge, CB3 9DE

January 2000
PRELIMINARY AND INCOMPLETE
(DO NOT QUOTE)

Abstract

In many published studies involving discrete choice a large class of models have been utilised without apparent consideration of both the statistical and behavioural relationships that exist between different model structures. In the case of multinomial models a commonly used behavioural interepretation is that the observed joint decision is made on the basis of utility maximisation over the choice set. We demonstrate that a more common model of discrete choice behaviour, the bivariate model, is a special case of the multinomial with rather restrictive implications for utility maximisation. We utilise the joint labour force participation and fertility decision problem as an example and subject a number of models to a series of nested and nonnested tests.

Acknowledgement. We are grateful to Paul Ruud and Christian Gourieroux for valuable comments and suggestions.

Keywords: discrete choice, multinomial response, bivariate models, labour supply, fertility, nonnested tests.

1 Introduction

Although microeconometric studies of discrete choice behaviour are now commonplace the majority of applications specify a single choice equation. In this respect any specification testing of an estimated model tends to focus on the adequacy of mean and distributional assumptions, with the implicit assumption that the decision process is independent of any other choices. The use of a joint estimation approach to modelling systems of discrete random variables has a rich history in both the econometrics and statistics literature. Zellner and Lee (1965) propose the use of generalised least squares as an extension of the *seemingly unrelated* regression model. However, although it is obvious that the rationale for such an approach follows from the use of estimators which are asymptotically more efficient than single equation methods, there are a number of additional issues which, we believe, are far more important than simply exploiting covariance across stochastic errors. Hensher and Johnson (1981) address this issue noting that "many choice structures cannot be adequately represented by a single choice equation either because the set of decisions are too complex ..., there is a hierarchical relationship between a set of interdependent choices, or a decision entails the simultaneous choice of a complex alternative" (p.65).

In this paper we consider a number of alternative models for modelling *systems* of discrete choice. We focus on the outcome of two decisions: the decision to have children and the labour force participation decision, which incorporates non-work, part-time and full-time options. In this context the work of Nakamura and Nakamura (1992) is instructive. In conducting an extensive overview of econometric models of labour supply and children, the authors propose a classification of approaches based upon whether, within a set of family decisions, variables can be categorised as either outcomes of a decision process or exogenous. Within this framework *predetermined* choice variables may be identified if they represent decisions made prior to the current time period. For example, if we assume that the labour supply decisions within a household are made *jointly* then the current earnings of the male would be treated as endogenous to female labour supply. Alternately, male earnings are exogenous in a household where labour supply decisions are made independently.

The concept of endogeneity is also important in terms of the manner in which we define the choice set. For example, in the context of modelling the participation decision and whether or not to have children, the discussion generally focusses upon the issue of how best to instrument for the endogenity of child status variables in a participation equation. However, this discussion implicitly assumes that the appropriate choice set is define over the binary (or multinomial) participation decision. As such the issue of whether the choice set should be constructed over the joint distribution of the two decisions, and how these approaches differ, has been ignored.

Unlike the majority of empirical studies in this area we subject a class of discrete choice models to a range of diagnostic (incomplete) and hypothesis tests. Despite a number of excellent texts by Greene (1993), Maddala (1983), edited series by Manski and McFadden (1981), there still seems to be some confusion amongst practitioners as to the precise nature of the underlying behavioural assumptions. In certain cases there exists a direct and transparent correspondence between the statistical and behavioural assumptions which underly competing models. This is demonstrated by the choice between the multinomial logit (MNL) and multinomial probit (MNP) models of discrete choice. In this instance the behavioural implications of the logistic versus multivariate normality assumption are manifest in a trade-off between a flexible error structure and one which imposes the independence of irrelevant alternative (IIA) assumption.

In other situations the behavioural implications of model choice are less well known. For example, although the bivariate probit model has been used extensively in discrete choice modelling, the behavioural assumptions that underly the model are, in general, not well known. Weeks and Orme (1999) demonstrated that the two binary decisions generating four mutually exclusive outcomes represent a nested and highly restrictive form of the more general multinomial probit model. In terms of the behaviorial implications, the authors have shown that the bivariate model implies a form of additive separability across the decisions, a restriction which is not shared by the more general multinomial model. Our approach is similar to the work of Hensher and Johnson (1981) who examine a number of alternative decision structures and analyse the structural relationship between decisions.

In Section 2 we present a brief overview of the different approaches taken in the economic literature to model the decision of women regarding participation and fertility. Section 3 examines the nature of the choice set for fertility and labour supply decisions and in doing so motivates a number of economic models of multiple decision making, examining the implications for the underlying utility calculus. The data is discussed in Section 4 and Section 5 presents our findings. Appendix I and II outline the structure of the econometric models.

2 Review of Literature

The application of economic reasoning to fertility or to fertility *and* participation has already a long tradition and has reached a high level of sophistication, with the estimation of complex dynamic models (see for example, Moffitt (1984) and Holtz and Miller (1988)). One of the predominant stylised facts that has emerged from the empirical literature is the existence of a strong negative correlation between the presence of young children in the household and female labour supply. Typically any measure of female labour supply (i.e. participation status or hours of work) is negatively correlated with any measure of young children. Influential early studies of labour supply which document this correlation include Mincer (1962), Cain (1966) and Bowen and Finegan (1969).

In the literature there are two broad approaches to estimating labour supply functions, taking account of fertility. The first is to estimate a reduced form model (hereafter RF) (see Browning (1992)). In this approach we do not include children as variables in the labour supply equation, although we may include variables that determine fertility. Given that one of the costs of having a child is the foregone earnings of the person caring for the child in the home, the wage rate was assigned a central role. This approach draws its inspiration from demand theory and the extension due to Becker (1960). Children are treated as a commodity and should not be included in other demand functions (for example the demand for leisure) anymore that we should put purchases of tea on the right hand side of a demand for coffee equation (Schultz (1978), Moffitt (1984), Rosenzweig and Wolpin (1980), Carliner, Robinson, and Tomes (1980).

An alternative approach is to include measures of fertility as conditioning variables with some allowance for endogeneity by instrumenting. A large number of investigators have followed this approach, usually without instrumenting. Moffitt (1984) notes that simple static models of labour supply which utilise the current stock of children as independent explanatory variables are suspect.The principal problems with this modelling strategy have been noted by Rosenzweig and Wolpin (1980). First, it is asserted that the only valid instrument for fertility variables in a labour supply equation is "the cost of increasing the number of children" (for example, birth and contraceptive costs). If this cost is not observed we cannot obtain consistent estimates of labour supply conditional on fertility. If, on the other hand, we do observe the cost of increasing family size, then labour supply equations conditional on this cost give all the information necessary to infer anything we want to know about the conditional labour supply equation. As such the standard approach is held to be either inconsistent or redundant.[1]

An excellent survey of applied econometric research on the effects of children on female labour supply is provided by Nakamura and Nakamura (1992). The principal focus of their review is the appropriate modelling framework with which to analyse *child status effects*. Whereas a number of other surveys have emphasised the impact of economic variables such as wage rates, the authors examine alternative approaches to incorporating children in models of labour supply. A unifying theme in their study is the question of whether fertility and labour supply represent joint decisions and related, the possible endogeneity of child status variables. In general, one would expect that within the context of a life cycle model of labour supply and fertility the decision to have children (and how many), and how many hours of work to supply are part of a joint decision problem and depend on the whole lifetime sequence of price and wages and on a variety of characteristics reflecting preferences. This approach would not justify the hypothesis of causation of one decision upon the other or of mutual causation; the two type of decisions are simultaneous, not only in the way of their timing but in the sense that they are the solution to a *common* constrained maximisation problem.[2] In static models the concept of endogeneity is also relevant when child status variables are included to reflect current demands on household resources, such as the level of child care. Mincer (1962) using data from the Bureau of Labour Statistics Survey of Consumer Expenditure provided evidence to support this argument and in particular demonstrated that labour supply and

[1] An additional problem with this approach is that even if we could obtain consistent estimates of the parameters of the labour supply equation conditional on fertility these would not be of any interest. The appropriate experiment for policy evaluation is held to be the effects on labour supply of changes in exogenous variables such as contraceptive costs, child care costs, or the price of "child intensive" goods.

[2] Note that it is still possible to speak of an exogenous effect of realised fertility upon participation or vice versa in the sense of deviations of the *realised* values from the *planned* ones which induce spill-over effects.

fertility choices are determined by the same set of economic variables. In this respect simultaneity is not a problem when fertility and labour supply decisions are viewed as joint consumer-demand choices.

Heckman and Willis (1976) note that much of the empirical literature on fertility has focussed upon single cross sections and as such has sought to explain the variation in the stock of children. The authors also point out that most static models of fertility have abstracted from the fact that the *fertility decision and realisations adds to and interacts with uncertainty surrounding other jointly determined household decisions.* Willis (1973) has also criticised the use of single equation labour supply and fertility models and advocated modelling the joint fertility-labour supply decision.

3 Systems of Discrete Choice

If we let $\mathbf{y}_i = (y_{i1}, ..., y_{iJ})'$ denote a $J \times 1$ vector representing discrete outcomes of J decisions for the *ith* individual, then if we abstract from the effect of any covariates, we may enumerate a number of possible approaches to model specification: (i) we could impose independence and model the marginal distribution of each binary response as a set of covariates i.e. estimate a single binary choice participation equation; (ii) specify a system of equations allowing for covariance across the set of discrete outcome, i.e. a seemingly unrelated regression model i.e. a bivariate choice model over participation and fertility decisions; (ii) allow for simultaneity amongst the elements of $\mathbf{y}_i$ i.e. estimate a simultaneous discrete choice model with possible instruments for child status variables in the participation equation; (iv) create a new variable based upon cartesian products of individual elements of $\mathbf{y}_i$ i.e. let $y_1 y_2$ denote the combination of the two discrete binary random variables representing labour supply and fertility.

One approach which has been adopted in the statistics literature is the use of a log-linear model. [*cite* Cox (1972) *on different approaches*] In this setting, parameters are interpreted in terms of the conditional distribution of a subset of variables given the other. Note that this approach is in contrast to the binary or multiple choice modelling where the focus is on the marginal distribution of each element of $\mathbf{y}_i$. Glonek and McCullagh (1986) in presenting a framework for examining a class of multivariate logistic models, compare the bivariate extension of the familiar logit model with the log-linear model. The authors emphasise that the principal disadvantage of the log-linear relative to the logistic model is that it lacks 'upward compatibility' (see McCullagh 1989)). However, in the bivariate extension of the univariate logistic representation, the marginal distribution of each variable has a univariate logistic representation[3]. [*restricted correlation in bivariate logistic*] In the class of log-linear models, the impact of any covariate on Y, depends upon the dimension of J. (See McCullagh and Nelder (1989)).

An approach suggested by Ashford and Sowden (1970), models correlated binary data by extending the binary probit model and modelling a system of binary equations or decisions. The authors, following Mantel (1966), are careful to point out the distinction between a model which recognises dependence over two distinct decisions which determine 4 mutually exclusive outcomes (iii in the above), and a single system providing four outcomes with an arbitrary dependence structure (iv in the above). (The multivariate and multinomial models extend this to consider dependence over h decisions and the 2^h mutually exclusive outcomes, respectively.) Apart from the comments by McFadden (1981), it appears that the precise nature of this distinction is not widely appreciated in economics. In the case of two equations the bivariate probit model has been used in many applications from voting behaviour (see Greene (1993)), to child-care and labour force participation (see Ribar (1992)). The extension of this model to higher dimensions encounters the curse of dimensionality as a result of flexible yet intractable multivariate normal distribution.

Below we consider a number of alternate representations of systems of discrete choice as applied to labour force and fertility choices.

3.1 Labour Force and Fertility Decisions

A decision is defined as the choice between a finite set of mutually exclusive outcomes. In developing notation we focus on two decisions y and q, and let $\mathbf{\Omega}$ and $\mathbf{\Theta}$ denote the respective choice sets. Within each choice set decisions are made on the basis of unobserved random variables. Letting y_j^* (q_p^*) denote the value (or utility) for the jth (pth) choice in $\mathbf{\Omega}$ ($\mathbf{\Theta}$), the mapping from the unobserved to revealed preference is generated according to the following decision rules:

[3]Note that the bivariate logistic model is restrictive in the sense that the correlation coefficient is bounded by the interval $|\rho| \leq 0.304$. SeeGumbel (1961).

$$y = \mathbf{1}(y_k^* = \max(y_j^*, \forall j \in \mathbf{\Omega})) \tag{1}$$

$$q = \mathbf{1}(q_t^* = \max(q_p^*, \forall p \in \mathbf{\Theta})), \tag{2}$$

where y and q are discrete random variables and $\mathbf{1}(\cdot)$ is the indicator function. Dependent upon the covariance structure both *within* and *across* $\mathbf{\Omega}$ and $\mathbf{\Theta}$ a number of alternative specifications are possible. For example, if both $\mathbf{\Omega}$ and $\mathbf{\Theta}$ contain just two elements, then dependent upon the distribution of the stochastic component of choice and the correlation across choice sets, we might choose between a bivariate logistic and bivariate probit specification.

Given (1) and (2) we may also enumerate four mutually exclusive outcomes for the pair (y, q). This is illustrated in Table 1 where $j = 1, ..., 4$ labels the outcomes, and v^* represents a value measure (or utility level) over the four states indexed by s.

Table 1

y	q	s	v^*
1	1	1	v_1^*
1	0	2	v_2^*
0	1	3	v_3^*
0	0	4	v_4^*

Although the above table represents an abstract system, there exist many examples from economics and bioassay where the relationship between two (or more) Bernoulli random variables and the associated states is important. For example, Cragg and Uhler (1970) consider a number of alternative frameworks for modelling systems of discrete choice behaviour. By focussing upon the demand for automobiles, the authors identify four options available to individuals who own at least one car: make no change, sell the car(s), sell *and* replace the car, purchase an additional car. A multinomial logit model, considering the joint decision over these alternatives is preferred to separate binary logit models.

In this paper we consider a class of discrete choice models based upon two decisions: labour force participation (P) and fertility (F). The choice set for each decision are given by

$$\begin{aligned} \mathbf{\Omega} &= \{NW,\ PT,\ FT\} \\ \mathbf{\Theta} &= \{C, \overline{C}\}, \end{aligned}$$

where NW, PT and FT are identifiers representing, respectively, individuals who do not work, part-time and full-time workers. C ($\overline{C}$) denote women with (without) children. If we take the Cartesian product of the elements in $\mathbf{\Omega}$ and $\mathbf{\Theta}$ and denote the resulting choice set by $\boldsymbol{\zeta}$, then the set of elements in $\boldsymbol{\zeta}$ may be written as

$$\boldsymbol{\zeta} = \{NW\_C,\ NW\_\bar{C},\ PT\_C,\ PT\_\bar{C},\ FT\_C,\ FT\_\bar{C}\},$$

where for example, the element PT_C denotes the *joint* decision of part-time work and children.

For issues related both to the economics of decision making *and* the tractability of econometric models, the way in which we analyse the decision process over the sets $\mathbf{\Omega}$ and $\mathbf{\Theta}$ is critical. If we consider the modelling of joint fertility and labour supply decisions there are a number of interesting research issues related to the manner in which we construct the choice set. Indeed, this component of model specification, namely how should the choice set be defined in such a way to reflect the mechanism of choice, has, with few exceptions, been neglected in the literature. In instances where we observe a time series of labour supply activity for each individual, then questions such as do labour supply decisions exhibit temporal dependence, and whether it is possible to distinguish this effect from the impact of any unobserved heterogeneity are of interest. However, if the analyst is confronted with the revealed preferences of a sample of individuals over a set of labour supply and fertility options for a *single* period, then an important but yet relatively unexplored issue is whether the utility from the two decisions is additively separable. As we will demonstrate below, the nature of the separability assumption is a direct consequence of how the choice set is defined. Browning (1992) considers the role of separability in the *intertemporal* decisions of households. In the case of intertemporal (weak) separability, demands are written as a function of current prices and current total expenditures. In the case of the consumption function and labour supply, the stronger assumption of intertemporal additivity of the utility function is made. In Browning's study based upon aggregate UK consumption data for the period 1972-1991, his results demonstrate that imposing additive intertemporal preferences results in considerable bias in the estimates of elasticities.

In this study we examine a number of alternative specifications which will enable us to determine whether additive separability of utility over labour supply and fertility behaviour is consistent with the data. Since our model is static we do not describe the inherently sequential nature of both the fertility and labour supply decisions but simply focus on constructing a *completed* fertility/labour supply econometric model[4] (see Moffitt (1984)). To focus attention we first describe the process which generates discrete observations on labour supply and fertility outcomes. We let $\mathbf{y}^p_{t^*|t}$ be a $t^* \times 1$, 0,1 vector of discrete labour supply decisions observed at time t. (We initially focus on the two labour market states but lose nothing in generality). Similarly we define $\tau^f_{t^*|t}$ as a $t^* \times 1$, 0,1 vector of fertility decisions such that $\tau^f_t = 1$ indicates an addition to the stock of children at time t.

In this study we have access to a single cross section of data. Letting s_{it} denote the age of the youngest child at time t, a binary fertility indicator $q^f_{it} = \mathbf{1}(0 < s_{it} < \zeta)$ simply records whether the woman has at least one child less than a threshold age, ζ. In this respect, a woman whose youngest child exceeds ζ is treated in the same way as a woman who does not have children. We choose to treat the fertility measure in this way given that our primary focus is the relationship between fertility and labour supply decisions. Thus our maintained hypothesis is that when $s_{it} > \zeta$ the stock of children for most women can be considered fixed, and any subsequent labour supply decisions are assumed to be independent of fertility.

Faced with data on a single cross-section of revealed preferences at time t, it is possible to explain *average* behaviour based upon the following joint sample frequencies

$$\begin{aligned}
\eta_1 &= \sum_{i=1}^{n} \mathbf{1}(y^p_{it} \cap q^f_{it}), \quad \eta_2 = \sum_{i=1}^{n} \mathbf{1}(y^p_{it} \cap (1 - q^f_{it})), \\
\eta_3 &= \sum_{i=1}^{n} \mathbf{1}(1 - y^p_{it}) \cap q^f_{it}), \quad \eta_4 = \sum_{i=1}^{n} \mathbf{1}((1 - y^p_{it}) \cap (1 - q^f_{it})),
\end{aligned} \tag{3}$$

where, for example $\mathbf{1}(y^p_{it} \cap q^f_{it})$, indicates that the *ith* individual works and has at least one child less than the threshold age. Note that since we only observe data for a single period we simply observe a stock of children. Although we will still refer to the fertility *decision*, the modelling framework does not allow us to model the decision of whether to have a child in a particular period, but rather

[4] see Heckman and Willis (1976) and Namboordi (1972) [*not in refs*] for a critique of econometric models of completed fertility.

to explain the variation over both *completed* fertility outcomes and participation decisions at a particular point in time.

Given the two binary variables y^p and q^f, the issue of separability is related to how we choose to model choice over the multinomial frequencies $\eta_j, j = 1, ..., 4$. For example, we could decide to focus upon separately modelling the marginal frequencies for labour force participation, using the aggregate frequencies $\eta_{12} = \eta_1 + \eta_2$ and $\eta_{34} = \eta_3 + \eta_4$, and fertility using $\eta_{13} = \eta_1 + \eta_3$ and $\eta_{24} = \eta_2 + \eta_4$. A variant of this approach is to model the marginal frequencies and allow for correlation across unobservables. In the case where the joint distribution of unobservables is distributed bivariate normal we have the bivariate probit model. Alternately we could model the labour supply decision *conditional* upon fertility, thereby treating the fertility variable analogously to other covariates.

Depending upon how we treat the two decisions - labour force participation and fertility - we may motivate a number of competing models of utility maximisation. Consider again Table 1 and assume that y denotes the outcome of the participation decision and q the fertility decision. We first examine the case where the joint decision, s, is made on the basis of utility maximisation, but by considering the *separate* utilities obtained from y over choice set $\mathbf{\Omega}$ and q over choice set $\mathbf{\Theta}$, rather than over the set ζ. To do this first write the unobserved utilities for the participation and fertility decision as a linear function of a set of covariates, $\mathbf{x}$. Specifically, write

$$\begin{aligned} y_p^* &= \mathbf{x}'\boldsymbol{\alpha}_p + u_p, \quad p = \text{NW}, \text{W} \\ q_f^* &= \mathbf{x}'\boldsymbol{\beta}_f + v_f, \quad f = \text{C}, \bar{\text{C}} \end{aligned} \tag{4}$$

where y^*_{NW} (y^*_{W}) is the utility derived from not working (working), q^*_{C} ($q^*_{\bar{\text{C}}}$) is the utility from having (not having) children, $\boldsymbol{\alpha}_p$ and $\boldsymbol{\beta}_f$ are vectors of unknown parameters and u and v are disturbance terms. Under the assumption of utility maximising behaviour on individual decisions, $p =$NW is observed if and only if $y^*_{\text{NW}} - y^*_{\text{W}} > 0$ and f=C *iff* $q^*_{\text{C}} - q^*_{\bar{\text{C}}} > 0$.

If we assume that the utilities derived from the joint outcome (y, q) denoted v_j^*, $j = 1, \ldots, 4$, are formed as $y_p^* + q_f^*$, then we may write

$$\begin{aligned} v_1^* &= y^*_{\text{NW}} + q^*_{\text{C}}, \qquad (5) \\ v_2^* &= y^*_{\text{NW}} + q^*_{\bar{\text{C}}}, \\ v_3^* &= y^*_{\text{W}} + q^*_{\text{C}}, \end{aligned}$$

$$v_4^* = y_{\text{W}}^* + q_{\bar{\text{C}}}^*.$$

This is a strong assumption; it says that the utilities, v_j^*, are *additively separable over the sets* $\boldsymbol{\Omega}$ *and* $\boldsymbol{\Theta}$. Note that, given (4), additive separability is not only sufficient, but it is also necessary for this representation of v_j^*.[5]

Alternately, if we consider utility maximisation over $\boldsymbol{\zeta}$, we might parameterise the utilities v_j^* as

$$v_j^* = \mathbf{x}'\boldsymbol{\gamma}_j + \varepsilon_j; \qquad j = 1, ..., 4 \tag{6}$$

in which the $\boldsymbol{\gamma}_j$ are vectors of unknown parameters. If $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)'$ is distributed multivariate normal with zero mean vector and covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{jl}\}$; i.e., $\boldsymbol{\varepsilon} \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$, then we have the multinomial probit model (MNP). The probabilities for each outcome may be written as

$$\begin{aligned}
p_1 = &\ \Pr\left(\eta_2 > -\mathbf{x}'\boldsymbol{\delta}_2 \cap \eta_3 > -\mathbf{x}'\boldsymbol{\delta}_3 \cap \eta_4 > -\mathbf{x}'\boldsymbol{\delta}_4\right), \\
p_2 = &\ \Pr\left(\eta_2 < -\mathbf{x}'\boldsymbol{\delta}_2 \cap (\eta_3 - \eta_2) > -\mathbf{x}'\left(\boldsymbol{\delta}_3 - \boldsymbol{\delta}_2\right) \cap (\eta_4 - \eta_2) > -\mathbf{x}'\left(\boldsymbol{\delta}_4 - \boldsymbol{\delta}_2\right)\right), \\
p_3 = &\ \Pr\left(\eta_3 < -\mathbf{x}'\boldsymbol{\delta}_3 \cap (\eta_3 - \eta_2) < -\mathbf{x}'\left(\boldsymbol{\delta}_3 - \boldsymbol{\delta}_2\right) \cap (\eta_4 - \eta_3) > -\mathbf{x}'\left(\boldsymbol{\delta}_4 - \boldsymbol{\delta}_3\right)\right), \\
p_4 = &\ \Pr\left(\eta_4 < -\mathbf{x}'\boldsymbol{\delta}_4 \cap (\eta_4 - \eta_2) < -\mathbf{x}'\left(\boldsymbol{\delta}_4 - \boldsymbol{\delta}_2\right) \cap (\eta_4 - \eta_3) < -\mathbf{x}'\left(\boldsymbol{\delta}_4 - \boldsymbol{\delta}_3\right)\right),
\end{aligned} \tag{7}$$

where $\boldsymbol{\delta}_j = \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_j$ and $\eta_j = \varepsilon_1 - \varepsilon_j$ will be determined by the joint distribution of $\boldsymbol{\eta} = (\eta_2, \eta_3, \eta_4)' \sim MVN(\mathbf{0}, \boldsymbol{\Omega})$.

3.2 The Relationship between MNP and BVP

The way in which the multinomial probit model formally nests the BVP model is described by the following proposition:

Proposition 1 *Let* (i) $\begin{pmatrix} \eta_2 \\ \eta_3 \end{pmatrix} \sim BVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$; (ii) $\eta_4 = \eta_2 + \eta_3$; (iii) $\boldsymbol{\delta}_2 = \boldsymbol{\alpha}_{\text{NW}} - \boldsymbol{\alpha}_{\text{W}}$, $\boldsymbol{\delta}_3 = \boldsymbol{\beta}_{\text{C}} - \boldsymbol{\beta}_{\bar{\text{C}}}$, $\boldsymbol{\delta}_4 = (\boldsymbol{\alpha}_{\text{NW}} - \boldsymbol{\alpha}_{\text{W}}) + \boldsymbol{\beta}_{\text{C}} - \boldsymbol{\beta}_{\bar{\text{C}}}$. *Then* $\pi_j = p_j$, $j = 1, \ldots, 4$.

The result is readily established and we demonstrate it only for p_1. Upon substitution of $(i) - (iii)$, in the expression for p_1, we can write the restricted probability p_1^r as

$$p_1^r = \Pr\left(\eta_2 > -\mathbf{x}'\boldsymbol{\alpha} \cap \eta_3 > -\mathbf{x}'\boldsymbol{\beta} \cap (\eta_2 + \eta_3) > -\mathbf{x}'\left(\boldsymbol{\alpha} + \boldsymbol{\beta}\right)\right), \tag{8}$$

[5] Of course, any linear combination of y_p^* and q_f^* will generate linear v_j^*, but arbitrary scale and location parameters in any such linear combination will not be identifiable from the bivariate data.

$$
\begin{aligned}
&= \Pr\left(\eta_2 > -\mathbf{x}'\boldsymbol{\alpha} \cap \eta_3 > -\mathbf{x}'\boldsymbol{\beta}\right), \\
&= \Phi_2\left(\mathbf{x}'\boldsymbol{\alpha}, \mathbf{x}'\boldsymbol{\beta}; \rho\right) = \pi_1.
\end{aligned}
$$

where $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{\text{NW}} - \boldsymbol{\alpha}_{\text{W}}$ and $\boldsymbol{\beta} = \boldsymbol{\beta}_{\text{C}} - \boldsymbol{\beta}_{\bar{\text{C}}}$. It is clear that the conditions identified in the above proposition lead to a constrained MNP model; i.e., the MNP model is algebraically equivalent to the BVP model if (6) is subject to parametric restrictions. Note that p_1 and π_1 denote, respectively, the probability of not working and having children based upon the multinomial probit and bivariate probit model. From (7) we see that given that $\boldsymbol{\zeta}$ has 4 states, p_1 is based upon the 3 pairwise differences generated by comparing the NW_C with all other states. However, if we allow individuals to maximise utility *separately* over the participation set, $\boldsymbol{\Omega}$, and the fertility set, $\boldsymbol{\Theta}$, then the relevant probability expression is given by π_1 in (8). In this instance $\mathbf{x}'\boldsymbol{\alpha}$ ($\mathbf{x}'\boldsymbol{\beta}$) represents the index function argument for participation (fertility) decision and ρ is the correlation coefficient. Comparing p_1 in (7) with p_1 in (8) we see the nature of the restrictions for the multinomial and bivariate model to coincide. Note that given additive utility, the 3 arguments of p_1^r in (7) are based upon the following comparisons

$$v_1^* - v_2^* > 0 \Rightarrow (y_{\text{NW}}^* + q_{\text{C}}^*) - (y_{\text{NW}}^* + q_{\bar{\text{C}}}^*) = q_{\text{C}}^* - q_{\bar{\text{C}}}^* > 0 \tag{9}$$

$$v_1^* - v_3^* > 0 \Rightarrow (y_{\text{NW}}^* + q_{\text{C}}^*) - (y_{\text{W}}^* + q_{\text{C}}^*) = y_{\text{NW}}^* - y_{\text{W}}^* > 0 \tag{10}$$

$$v_1^* - v_4^* > 0 \Rightarrow y_{\text{NW}}^* + q_{\text{C}}^*) - (y_{\text{W}}^* + q_{\bar{\text{C}}}^*) > 0 \tag{11}$$

$$\Rightarrow (y_{\text{NW}}^* - y_{\text{W}}^*) + (q_{\text{C}}^* - q_{\bar{\text{C}}}^*) > 0. \tag{12}$$

where (9) represents the fertility decision, (10) the participation decision, and since (11) represents the sum of (9) and (10), this condition is redundant.

4 A Behavioural Interpretation

Below we provide a brief overview of the random utility model. Let U_{ij} denote the utility of the jth alternative for the ith individual. In addition we follow the standard approach and decompose utility into a deterministic (V_{ij}) and stochastic component[6] (ε_{ij}), giving

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{13}$$

[6]Note: the concept of random utility has two interpretations. First, the stochastic component of utility can be interpreted as those determinants of choice which are unobserved by the analyst. In this respect individuals maximise utility in a deterministic fashion. Alternately

where ε_{ij} reflects individual specific variation in tastes. Further we let V_{ij} depend on both attribute specific and individual invariant characteristics, such that

$$V_{ij} = \alpha_j + x_{jk}\beta_k + v_t\theta_{jt} \tag{14}$$

where $x = \{x_{jk}\}$ is a $J \times K$ matrix of non-stochastic components of utility, whose rows contain the K *alternative specific* attributes of alternative $j; v = \{v_t\}$ is a $T \times 1$ vector of *alternative invariant* individual characteristics; $\alpha = \{\alpha_j\}, \beta = \{\beta_k\}, \theta = \{\theta_{jt}\}$ are respectively $J \times 1$, $K \times 1$, and $J \times T$ arrays of unknown parameters.

We assume that individuals choose the alternative that generates the largest utility. Alternative l is chosen iff

$$U_{il} > U_{ij} \qquad \forall_j \neq l \in \zeta, \tag{15}$$

where ζ denotes the choice set. Moving from (15) to a statement about the probability that l is chosen involves the imposition of a distributional assumption on the pairwise error differences of the stochastic components. For example, given (13) an alternative way of writing (15) is

$$V_{il} - V_{ij} > \varepsilon_{ij} - \varepsilon_{il} \qquad \forall_j \neq l \in \zeta$$

or

$$P_{il} = \Pr(\varepsilon_{ij} < \varepsilon_{il}) < (V_{il} - V_{ij})$$

4.1 A Comparison of Models

The empirical analysis evaluates the extent to which a class of discrete choice models are consistent with the observed data on labour force participation (D_2) and fertility (D_1) decisions. We consider two broad approaches:

I. Utility maximisation is performed separately over the two choice sets $\mathbf{\Omega}$ and $\mathbf{\Theta}$. Weeks and Orme (1999) demonstrated that if we allow a non-zero parameter, ρ, to allow for correlation *across* the two decisions D_1 and D_2, then the set of utilities given by (5) are consistent with the specification of

we could postulate that even if all attributes of alternatives and characteristics of individuals were fully observed, that utility maximisation is random. We follow the literature and interpret random utility according to the former explanation.

a bivariate discrete choice model. Dependent upon the assumed joint distribution of the stochastic components, possible models include the bivariate probit and bivariate logistic discrete choice models.

II. Utility maximisation is over the set $\boldsymbol{\zeta}$ and utility is *not* additively separable over $\boldsymbol{\Omega}$ and $\boldsymbol{\Theta}$. Under this scenario we have the *multinomial* response model.

Although we believe that in many situations the distinction between models I and II is not appreciated, considerable use has been made of the concept of separability in demand structures. For example, a hierarchical choice structure would facilitate a reduction in the dimensionality of the calculus required by optimising agents. Thus, a consumer faced with the decision to choose a bundle of goods from a large number of possibilities, say α, might partition α into distinct categories and perform separate maximisation on each category. Related problems occur within the context of lifetime decision problems, with the decision maker separating decisions on current from future choices.

In Figure 1 we depict two different decision structures which characterise the labour force (D_2) and fertility decision (D_1). Figure 1(a) presents the bivariate model with ρ denoting the correlation coefficient of the bivariate normal distribution. Note that here we let the labour force participation decision have three outcomes by differentiating between part-time and full-time work. Figure 1(b) presents the same two decisions but is based on the assumption that individuals maximise utility over a *single* choice set which is the Cartesian product of the choice set in D_1 and D_2. In this context the appropriate statistical model is multinomial.

In this paper we utilise a series of nested and nonnested hypothesis tests to evaluate the empirical adequacy of a number of discrete choice models. The models we examine are:

i. bivariate probit (2×2), (3×2)

ii. bivariate logit (2×2), (3×2)

iii. multinomial logit 4 state, 6 state

iv. multinomial probit 4 state.

5 Data

The data set used in this study is the British Household Panel Survey (BHPS). This survey is carried out by the ESRC Research Centre on Micro-social change at the University of Essex. The BHPS was designed as an annual survey of each adult (16+) member of a nationally representative sample of more than 5,000 households, making a total of approximately 10,000 individual interviews. The same individuals were re-interviewed in successive waves and, if they split off from original households, all adult members of their new households were also interviewed. At the moment there are 7 available waves from 1990 to 1996.

The individual questionnaire covers the following topics: neighbourhood, individual demographics, residential mobility, health and caring, current employment and earnings, employment changes over the past year, lifetime childbirth, marital and relationship history (Wave Two only), employment status history (Wave Two only), values and opinions, household finances and organization.[7] In this application we use the 5th wave of the BHPS where individuals were asked to report information about their situation in September 1994. Given that the purpose of the paper is to study participation and decision choices of women, we selected all the households with a married and cohabiting women, aged between 18 and 60. We assume that this life span covers the period in which women make their choices about working and about fertility. In other words we assume that women younger than 18 and older than 60 do not have to decide whether to work or whether to have children. Based upon these criteria the resulting sample contains 2502 observations. We choose to exclude women under 18 because we assume that their decision process about having children and participate to the labout market is characterised by different variables respect to the women over 18. 18 years old is the age where most individuals have completed their high school education.

Table 1 presents a description of the variables used in the analysis and Table 2 presents summary statistics. In this sample 34.53% of the women do not work, 31.10 % work part time (less than 30 hours per week) and 34.37 work full time. Fertility is a binary variable equal to 1 if the woman has at least one child between 0 and 15 years old, it is equal to 0 othewise. Among the variables related to the

[7] For a full description of the survey see BHPS User Manual: volume A.

budget set, we assign a prominent role to the wage of the woman. This choice is due to the fact that woman's wage enter both the opportunity cost of children and the net value of participation. However, the use of the wife's wage as an explanatory variable poses two problems. First, the wage is observed only for working women. Second the wage is likely to be correlated with unobserved variables affecting both fertility and participation decisions.

Here we adopt the standard solution to these problems (see for instance Moffitt (1984)), which consists in estimating a wage equation on the working subsample and correcting for selection bias using a procedure suggested by Heckman (1976). In this way the simultaneity problem (2) is solved with a 2SLS technique, and the unobservability problem (1) is solved by imputing the predicted wage to every woman both working and not working.

The woman's gross wage rate (dependent variable in the wage equation) is calculated using the usual gross pay per month, the number of hours normally worked per week and the number of overtime hours normally worked per week. The explanatory variables used in this equation are the dummies for education, experience and experience squared. The results are presented in table 3.

The predicted net hourly wage is calculated from the predicted gross wage rate applying the rates and bands for the 1993/94 and the National Insurance Contributions.[8] In order to be able to calculate the net hourly wage from the predicted gross hourly wage, it is necessary to calculate the gross annual labour income of the women in the sample. Since we do not observe hours of work for women who do not work, we estimated a regression for women's hours of work. The estimated parameters are then used to calculate predicted hours of work for all the women in the sample. The equation for hours of work is presented in table 4 and the explanatory variables used are the dummies for education, experience, experience squared and the children dummies.

[8] The tax rates used to calculate the predicted net wage are the following: 20% until £ 2,500, 25% for the next £21,200, and 40% for income above 23,700. The personal allowance used is equal £3,445 while the married allowance is equal to £ 1,720. The National insurance contributions used are as follow: 2% until £56 per week, 9% for the next £363 per week, 0 thereafter.

6 Results

In tables 5 and 6 we present results for the (2×2) and (3×2) bivariate models. In table 5 the dependent variables are the binary indicators for fertility and the binary indicator of labour force participation. Note that for the fertility equation parameter estimators are based upon the effects of covariates relative to the non fertility state $\bar{C}$; in the participation equation the reference group is non-workers (NW). For both models the bivariate logit and probit equation parameter estimates have the correct signs and are consistent with previous studies. In table 6 we extend the bivariate model by differentiating between part-time and full-time workers. Here we focus upon the effect of the predicted net wage of the woman. As expected this variable enters the fertility decision with a negative sign given the change in the opportunity cost of having children. However, the effect of an increase in the wage has quite a different order of magnitude effect in the part-time and full-time participation equations.

In all discrete choice models, and particularly in the case of multinomial models, parameter estimates are often difficult to interpret. Although this is also true in this study, the results in table 7 present the parameter estimates by using a multinomial logit model to estimate a 4 state labour supply/fertility decision based upon the convolution of the two choice sets which underly the results in table 5. In this case we use the NW_C state - non-working/no children as the referent alternative. Notable results include the fact that, as expected, an increase in the predicted wage results in an increase in labour supply but that this effect is more pronounced for women with children than those without (compare 0.309 with 0.196). Similar differences for these two groups are observed when we examine the impact of age. Compare the parameter estimator for women with children (W_C) with women without children (W_$\bar{\text{C}}$), again relative for the NW_$\bar{\text{C}}$ state.

In Table 7 we present the multinomial logit counterpart of the results in table 5. Again we use the NW_$\bar{\text{C}}$ as the referent alternative. Despite an obvious proliferation of parameter estimates, the results are consistent with expectations, and allow a more disaggregate perspective on joint decision making.

Table 8 reports the elasticties of the probabilities of participation and having children with respect to the female wage for the (2 x 2) models. The elasticity of the probability of participation to the labour market are, as expected, positive, while the elasticities of fertility are negative. One interpretation is that the wage represents an estimate of the opportunity cost of children. An increase in the female wage decreases the probability of having children. However, this effect is much smaller (in absolute terms) than the positive effect on participation: this result is consistent with estimates on other data sets.

We have also estimated the models presented for different age-selected samples: we have estimated the (2x2) bivariate logit and probit and the four states multinomial logit on samples of women under 50 years old and under 40 years old.Table 8 reports elasticities for these samples as well. The results show that the elasticities of participation with respect to female wage are decreasing if we exclude older women in the sample (the only exception.is the multinomial logit for the sample under 40). As far as the elasticities of having children are concerned, their absolute values increase when we exclude older women in the sample (as we would expect)

In the models that differentiate between full and part time work (table10), an interesting result comes out. The elasticities of full time participation are smaller in absolute value than the part time, and negative. This result is consistent with other estimates and other data for the UK (see, for example, Duncan and Weeks (1997)). One possible interpretation of the relative rigidity of full time female labour supply with respect to part time is that women who work full time might possibly be both more educated and motivated. At the margin, we have a backward bending labour supply function where the income effect is predominant. Part time workers, on the contrary, have higher elasticity given that their participation is highly sensitive to changes in wages.

6.1 Non-Nested Tests

Table 11 presents the results of an application of a two-sided version of Cox's non-nested test (see Cox (1961) and Cox (1962)) to the choice between a number of discrete choice models of labour supply and fertility.[9] Given the large sample size we utilise White (1982) estimator for the pseudo-true value in the construction of the numerator of the test statistic. Weeks (1996) showed that in large samples the size properties of this computationally convenient method for constructing the

[9] For an extensive overview of the use of nonnested testing to evaluate empirical adequacy see Pesaran and Weeks (1999).

numerator is comparable to the simulation estimator. We utilise three variants of the denominator based upon using the observed Hessian (H), an outer product of the gradient (OPG) estimator (O), and an estimator which ignores variation due to parameter estimation under the null (X) (see Pesaran and Pesaran (1995). Therefore, the test statistic C_{WH} is the Cox test statistic using White's method for the numerator and the observed Hessian (under the null) for the denominator. C_{SH}^{30} denotes the Cox test.

In interpreting the results we first note that in the case of nonnested models there is no natural null model. As such in testing two models, say H_f and H_g, we need to conduct two tests based upon interchanging the null and alternative hypothesis. The first row of Table 11 presents the three null models: multinomial logit (MNL), bivariate logit (Biv-L) and bivariate probit (Biv-P). In the second row H_g denotes the alternative models. For example, we test H_f: MNL against H_g:Biv-L and H_g' :Biv-P. The first block of results (I) presents test statistics for the $(2\times 2)/4$ choice models and in (II) we present results for the $(3\times 2)/6$ choice models.[10] Although we emphasise that our findings are preliminary, we report an internal consistent set of results. Namely, in all cases the bivariate models are rejected and we *fail to reject* the multinomial counterparts.

6.2 A Score Test

Since the parametric restrictions, described in Section 3.1, imply a singular covariance matrix in the MNP model, the most natural likelihood ratio procedure is not strictly available due to the problem of testing on the boundary of the parameter space. A score test procedure is therefore outlined, whose asymptotic validity is unaffected by such a problem. First we define an appropriate $(m\times 1)$ vector of unrestricted parameters $\boldsymbol{\theta}' = (\boldsymbol{\delta}', \boldsymbol{\sigma}')$ for the MNP model, which accommodates the BVP model as a special case:

$$\boldsymbol{\delta}' = (\boldsymbol{\delta}_2', \boldsymbol{\delta}_3', \boldsymbol{\delta}_4'), \qquad \boldsymbol{\sigma}' = (\sigma_{23}, \sigma_{14}, \sigma_{24}, \sigma_{34}, \sigma_{44}),$$

where $\boldsymbol{\Sigma} = \{\sigma_{jl}\}, j, l = 1, \ldots, 4$, and $m = 5+3k$. Conditional on $\mathbf{x}$, let the relevant four probabilities, (7), be expressed as functions of $\boldsymbol{\theta}$; i.e., $p_k(\boldsymbol{\theta}|\mathbf{x})$. Then based on N independent realisations of the indicator c_j, denoted c_{ij}, $i = 1, \ldots, N$, the

[10]Under H_f the suitably standardised Cox test statistic is asymptotically distributed as N(0,1). Given a 5% significance level the rejection region is therefore $> |1.96|$

log-likelihood is $\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{j=1}^{4} c_{ij} \ln(p_j(\boldsymbol{\theta}|\mathbf{x}_i))$. From this the $(m \times 1)$ score vector and $(m \times m)$ Hessian matrix are

$$\begin{aligned} \mathbf{g}(\boldsymbol{\theta}) &= \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{N} \sum_{j=1}^{4} c_{ij} \frac{\partial \ln(p_j(\boldsymbol{\theta}|\mathbf{x}_i))}{\partial \boldsymbol{\theta}}, \\ \mathbf{H}(\boldsymbol{\theta}) &= \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \sum_{i=1}^{N} \sum_{j=1}^{4} c_{ij} \frac{\partial^2 \ln(p_j(\boldsymbol{\theta}|\mathbf{x}_i))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}. \end{aligned}$$

Now let $\left(\tilde{\boldsymbol{\alpha}}', \tilde{\boldsymbol{\beta}}', \tilde{\rho}\right)$ denote (restricted) maximum likelihood estimates of the BVP model, (**??**), and define accordingly the restricted MNP maximum likelihood estimates as

$$\begin{aligned} \tilde{\boldsymbol{\delta}}' &= \left(\tilde{\boldsymbol{\alpha}}', \, \tilde{\boldsymbol{\beta}}', \, \tilde{\boldsymbol{\alpha}}' + \tilde{\boldsymbol{\beta}}'\right) \\ \tilde{\boldsymbol{\sigma}}' &= \left(\tilde{\rho}, \, 0, \, \tfrac{1}{2} + \tilde{\rho}, \, \tfrac{1}{2} + \tilde{\rho}, \, 1 + 2\tilde{\rho}\right). \end{aligned}$$

An asymptotically valid test of the implied $q = 4 + k$ parameter restrictions can be based on the score test statistic given by $\mathcal{S} = N^{-1}\mathbf{g}(\tilde{\boldsymbol{\theta}})' \left\{\mathbf{V}\left(\tilde{\boldsymbol{\theta}}\right)\right\}^{-1} \mathbf{g}(\tilde{\boldsymbol{\theta}})$, where $\mathbf{V}\left(\tilde{\boldsymbol{\theta}}\right)$ is any consistent estimator for the average information matrix, under the assumption that the restrictions under test are valid; e.g., $\mathbf{V}\left(\tilde{\boldsymbol{\theta}}\right) = -\frac{1}{N}\mathbf{H}\left(\tilde{\boldsymbol{\theta}}\right)$. In large samples, $\mathcal{S}$ is distributed as a chi-square random variable with q (number of restrictions) degrees of freedom when the parametric restriction imposed on the MNP model are correct. Significantly large values of $\mathcal{S}$ would be provide evidence against the BVP model and, as a consequence, it would also suggest that the BVP specification is inconsistent with utility maximisation over the implied four possible outcomes.

In the application of this test to the data we reject the restricted BVP model in favour of the more general multinomial probit model.

7 Conclusion

In much of the literature on discrete choice modelling of labour supply and fertility, there has been a tendency to ignore a fundamental component of model specification: namely, for a set of decisions, what is the appropriate form of the choice set, and related, what are the implications of different choice sets. In this paper we have examined this issue and demonstrated that the behaviourial consequences which follow from the use of different models. Based upon the application of a number of tests, our findings suggest that the more general multinomial models, which follow from the combination of decisions, are a better representation of the data. Although we believe that this finding is important, we offer a word of caution. As we extend this approach, by defining choice sets over an increasing number of decisions, there follow dimensionality problems, both in terms for numerical analysis and the implied calculus of decision makers faced with the choice over a large number of alternatives.

Appendix I: Econometric Models

Below we set out the structure of the bivariate and multinomial choice models.

Bivariate Probit

We begin by examining the structure of the model underlying the two state participation equation and a two state fertility equation. We refer to this as the 2×2 bivariate model. As in Section 3 we let y_p^* and q_f^* represent the latent variables which underly the participation and fertility decisions. We assume that y_p^* and q_f^* are generated according to the parametric specification given by

$$
\begin{aligned}
y_p^* &= \mathbf{x}'\boldsymbol{\alpha}_p + u^p \quad p = 0,1 \\
q_f^* &= \mathbf{z}'\boldsymbol{\beta}_f + v^f, \quad f = 0,1,
\end{aligned} \tag{16}
$$

where $y_p^* = y_{\mathrm{NW}}^* - y_{\mathrm{W}}^*$ and $q_f^* = q_{\mathrm{C}}^* - q_{\bar{\mathrm{C}}}^*$. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown parameter vectors, $\mathbf{x}$ is a $k \times 1$ vector of regressors and the joint distribution of u^p and v^f is *bivariate* normal. If we denote the vector of stochastic components by $\boldsymbol{\varepsilon}^* = (u_0^p,\ u_1^p,\ v_0^f,\ v_1^f)'$, where u_0^p, u_1^p denotes the disturbance term for the non-participation and participation working equation, and v_0^f, v_1^f denotes the disturbance term for the no children and children states, then the lower diagonal of the covariance matrix may be written as

$$
\sum = \begin{bmatrix}
\underbrace{\begin{matrix} \sigma_0^{2p} & \\ \sigma_{10}^p & \sigma_1^{2p} \end{matrix}}_{\mathbf{A}} & \\
\underbrace{\begin{matrix} \sigma_{00}^{fp} & \sigma_{01}^{fp} \\ \sigma_{10}^{fp} & \sigma_{11}^{fp} \end{matrix}}_{\mathbf{C}} & \underbrace{\begin{matrix} \sigma_0^{2f} & \\ \sigma_{10}^f & \sigma_1^{2f} \end{matrix}}_{\mathbf{B}}
\end{bmatrix}, \tag{17}
$$

where submatrices $\mathbf{A} = cov(u_0^p, u_1^p)$, and $\mathbf{B} = cov(v_0^f, v_1^f)$ respectively denote covariance matrices for the participation and fertility decisions. The matrix $\mathbf{C}$, with elements $\sigma_{jk}^{fp}\ \forall j, k = 0, 1$, denotes the covariance between the j and kth stochastic terms for, respectively, the fertility and participation decision. As such we can think of $\mathbf{A}$ and $\mathbf{B}$ as representing the contemporaneous covariance across states *within* the participation and fertility decision. $\mathbf{C}$ represents the contemporaneous covariance *across* the two decisions.

For both the participation and fertility decision we introduce a normalisation. For the participation decision we evaluate choice relative to the not working state, and for the fertility decision we evaluate choice relative to the no children state. We write the subsequent vector of differenced disturbance terms as $\boldsymbol{\varepsilon}_d^* = (u_{10}^p, v_{10}^f)$, where, for example, $u_{10}^p = u_1^p - u_0^p$. Note that $\boldsymbol{\varepsilon}_d^*$ includes other non-participant differences as exact linear combinations.[11] The covariance matrix of $\boldsymbol{\varepsilon}_d^*$ may be written as

$$\Omega = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{18}$$

where ρ is the correlation coefficient.

If we now extend the participation equation to three states then we have $\boldsymbol{\varepsilon}^* = (u_0^p, u_1^p, u_2^p, v_0^f, v_1^f)'$, where in this instance u_1^p and u_2^p denote the disturbance terms for the part-time and full-time states. The covariance matrix for $\boldsymbol{\varepsilon}^*$ may then be written as

$$\sum = \begin{bmatrix} \underbrace{\begin{matrix} \sigma_0^{2p} & & \\ \sigma_{10}^p & \sigma_1^{2p} & \\ \sigma_{20}^p & \sigma_{21}^p & \sigma_2^{2p} \end{matrix}}_{\mathbf{A}} & \\ \underbrace{\begin{matrix} \sigma_{00}^{fp} & \sigma_{01}^{fp} & \sigma_{02}^{fp} \\ \sigma_{10}^{fp} & \sigma_{11}^{fp} & \sigma_{12}^{fp} \end{matrix}}_{\mathbf{C}} & \underbrace{\begin{matrix} \sigma_0^{2f} & \\ \sigma_{10}^f & \sigma_1^{2f} \end{matrix}}_{\mathbf{B}} \end{bmatrix} \tag{19}$$

where the submatrices $\mathbf{A} = cov(u_0^p, u_1^p, u_2^p)$ and $\mathbf{B} = cov(v_0^f, v_1^f)$. $\mathbf{C}$ is defined as before, and denotes the covariance terms *across* the 3 state participation decision and the 2 state fertility decision. By introducing the same normalisation as above we may write the vector of differenced disturbance terms as $\boldsymbol{\varepsilon}_d^* = (u_{10}^p, u_{20}^p, v_{10}^f)'$, where

$$e_d^* \sim TVN \begin{pmatrix} \begin{matrix} 0 \\ 0 \\ 0 \end{matrix}, & \begin{bmatrix} \sigma_{10}^{2p} & & \\ \sigma_{20,10}^p & \sigma_{20}^{2p} & \\ \sigma_{10,10}^{fp} & \sigma_{10,20}^{fp} & \sigma_{10}^{2f} \end{bmatrix} \end{pmatrix} \tag{20}$$

If we let $\mathbf{x}'\boldsymbol{\alpha}_{0j}$, $j = 1, 2$, denote the index function for the participation equation and $\mathbf{z}'\boldsymbol{\beta}_{01}$ denote the index function for the fertility equation, the *marginal* probabilities for each of the two decisions are given by

[11] i.e. $u_{21}^p = u_{20}^p - u_{10}^p$.

$$
\begin{aligned}
P_l &= \Pr(P = l) = \mathbf{x}'\boldsymbol{\alpha}_{lj} > \varepsilon_{jl} \quad \forall_j \neq l \\
F_k &= \Pr(F = k) = \mathbf{x}'\boldsymbol{\beta}_{ks} > \varepsilon_{sk} \quad \forall_s \neq k,
\end{aligned} \tag{21}
$$

where, for example, the probability of non-participation, $\Pr(P = 0)$, is given by

$$
\Phi^2(\mathbf{x}'\boldsymbol{\alpha}_{01}, \mathbf{x}'\boldsymbol{\alpha}_{02}, \rho) = \int_{-\infty}^{\mathbf{x}'\boldsymbol{\alpha}_{01}} \int_{-\infty}^{\mathbf{x}'\boldsymbol{\alpha}_{02}} \phi^2(u_{10}^p, u_{20}^p, \rho) du_{20}^p, du_{10}^p,
$$

where $\Phi^2(.)$ and $\phi^2(.)$ respectively denote the bivariate normal cumulative distribution and bivariate normal probability density function with correlation ρ.

We write the *joint* probabilities for the six outcomes defined over the combined participation and fertility choice space as

$$
\begin{aligned}
P_0F_0 &= \Pr(P = 0 \cap F = 0) = \Phi^3(\mathbf{x}'\boldsymbol{\alpha}_{01},\ \mathbf{x}'\boldsymbol{\alpha}_{02},\ \mathbf{z}'\boldsymbol{\beta}_{01}, \rho_1, \rho_2, \rho_3) \\
P_0F_1 &= \Pr(P = 0 \cap F = 1) = \Phi^2(\mathbf{x}'\boldsymbol{\alpha}_{01},\ \mathbf{x}'\boldsymbol{\alpha}_{02},\ \rho_1) - P_0F_0 \\
P_1F_0 &= \Pr(P = 1 \cap F = 0) = \Phi^3(\mathbf{x}'\boldsymbol{\alpha}_{10},\ \mathbf{x}'\boldsymbol{\alpha}_{12},\ \mathbf{z}'\boldsymbol{\beta}_{01}, -\rho_1, \rho_2, -\rho_3) \\
P_1F_1 &= \Pr(P = 1 \cap F = 1) = \Phi^2(\mathbf{x}'\boldsymbol{\alpha}_{10},\ \mathbf{x}'\boldsymbol{\alpha}_{12}, -\rho_1) - P_1F_0 \\
P_2F_0 &= \Pr(P = 2 \cap F = 0) = \Phi(\mathbf{z}'\boldsymbol{\beta}_{01}) - P_0F_0 - P_1F_0 \\
P_2F_1 &= \Pr(P = 2 \cap F = 1) = 1 - P_0F_0 - P_0F_1 - P_1F_0 - P_1F_1 - P_2F_0 \\
&= 1 - F_0 - P_0F_1 - P_1F_1
\end{aligned} \tag{22}
$$

where $\Phi^3(a, b, c; \rho_1, \rho_2, \rho_3)$ denotes the trivariate normal distribution evaluated at the point (a, b, c), $\Phi^2(a,\ b;\ p)$ denotes a similarly defined bivariate normal distribution, and $\Phi(a)$ is the standard normal distribution evaluated at a. $\rho_1 = corr\ (u_{10}^p,\ v_{20}^f)$, $\rho_2 = corr\ (u_{20}^p,\ v_{10}^f)$, $\rho_3 = corr\ (u_{10}^p,\ u_{10}^f)$.

The MNP Model

For the MNP model, and focussing on the six state model, and using the same notation as in (6) we write the vector of latent variables $\mathbf{v}^*$ as

$$
v_j^* = \mathbf{x}'\boldsymbol{\gamma}_j + \varepsilon_j; \quad j = 1, ..., 6
$$

where $\boldsymbol{\gamma}_j$ are vectors of unknown parameters and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6)'$ is distributed multivariate normal with zero mean and covariance matrix $\boldsymbol{\Sigma} = \{\sigma_j\}$. Letting $j = 1$ index the non-work/no children state, this probability may be written as

$$\begin{aligned} P_0F_0 &= \Pr(v_{12}^* > 0 \cap v_{13}^* > 0 \cap v_{14}^* > 0 \cap v_{15}^* \cap v_{16}^* > 0) \qquad (23) \\ &= \Pr(\varepsilon_{21} < \mathbf{x}_{12}'\boldsymbol{\gamma} \ \cap \varepsilon_{31} < \mathbf{x}'\boldsymbol{\gamma}_{13} \ \cap \varepsilon_{41} < \mathbf{x}_{14}'\boldsymbol{\gamma} \ \cap \varepsilon_{51} < \mathbf{x}'\boldsymbol{\gamma}_{15} \ \cap \varepsilon_{61} < \mathbf{x}'\boldsymbol{\gamma}_{16}) \end{aligned}$$

where $\varepsilon_{j1} = \varepsilon_j - \varepsilon_1$ and $\boldsymbol{\gamma}_{12} = \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2$. Notice that (23) is given by a five-fold integral whereas the same probability underlying the BVP is given by a trivariate integral. For the BVP model the probability of no children and not working is given by

$$\begin{aligned} P_0F_0 &= \Pr(P = 0, F = 0) \\ &= \Phi^3(\mathbf{x}_{01}'\boldsymbol{\alpha},\ \mathbf{x}_{02}'\boldsymbol{\alpha},\ \mathbf{z}'\boldsymbol{\beta}_{01}, \rho_1, \rho_2, \rho_3) \end{aligned}$$

References

Ashford, J., and R. Sowden (1970): "Multivariate Probit Analysis," *Biometrics*, 26, 535–546.

Becker, G. (1960): "An Economic Analysis of Fertility," *Demographic and Economic Change in Developed Countries*, New York: NBER.

Bowen, W. G., and A. Finegan, Aldrich (1969): *The Economics of Labour Force Participation*. Princeton University Press, Princeton, N.J.

Browning, M. (1992): "Children and Household Economic Behavior," *Journal of Economic Literature*, XXX, 1434–1475.

Cain, G. G. (1966): *The Labor Force Participation of Married Women*. University of Chicago Press, Chicago.

Carliner, G., C. Robinson, and N. Tomes (1980): "Female Labour Supply and Fertility in Canada," *Canadian Journal of Economics*, XIII(1), 46–64.

Cox, D. (1961): "Tests of Separate Families of Hypothesis," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability.*

——— (1962): "Further Results on Tests of Separate Families of Hypotheses," *Journal of Royal Statistical Society*, B24, 406–424.

Cox, D. R. (1972): "The Analysis of Multivariate Binary Data," *Applied Statistics*, vol. 21, 113–126.

Cragg, J., and R. Uhler (1970): "The Demand for Automobiles," *Canadian Journal of Economics*, 3(3), 386–406.

Duncan, A., and M. Weeks (1997): "Non-Nested Models of Labour Supply with Discrete Choices," Department of Applied Economics, University of Cambridge (submitted to Journal of Applied Econometrics).

Greene, W. H. (1993): *Econometric Analysis*. Macmillan Publisher Company, 2nd edn.

Gumbel, E. (1961): "Bivariate Logistic Distributions," *American Statistical Association Journal.*

HECKMAN, J., AND R. WILLIS (1976): "Estimation of a Stochastic Model of Reproduction: An Econometric Approach," in *Household Production and Consumption*, ed. by N. Terleckyj. Columbia University Press, New York.

HENSHER, D. A., AND L. W. JOHNSON (1981): *Applied Discrete-Choice Modelling*. John Wiley and Sons, New York.

HOLTZ, J., AND A. MILLER (1988): "An Empirical Analysis of Life Cycle Fertility and Female Labour Supply," *Econometrica*, 56(1), 91–118.

MADDALA, G. (1983): *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

MANTEL, N. (1966): "Models for Complex Contingency Tables and Polychotomous Dosage Response Curves," *Biometrics*, 22, 83–95.

MCCULLAGH, P., AND J. NELDER (1989): *Generalized Linear Models (2nd Ed.)*. Chapman and Hall, London.

MCFADDEN, D. (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski, and D. McFadden, pp. 198–272. Cambridge MA:MIT Press.

MINCER, J. (1962): "Labor Force Participation of Married Women: A Study of Labor Supply," in *Aspects of Labor Economics*, ed. by L. H. Gregg, pp. 63–97. Princeton University Press, Princeton, N.J.

MOFFITT, R. (1984): "Profiles of Fertility, Labour Supply and Wages of Married Women: A Complete Life-Cycle Model," *Review of Economic Studies*, 51(2), 263–78.

NAKAMURA, A., AND M. NAKAMURA (1992): "The Econometrics of Female Labor Supply and Children," *Econometric Reviews*, 11(1).

PESARAN, M. H., AND B. PESARAN (1995): "A Non-Nested Test of Level Differences Versus Log-Differenced Stationary Models," *Econometric Reviews*, 14(2), 213–27.

PESARAN, M. H., AND M. WEEKS (1999): "Non-Nested Hypothesis Tests," in *Theoretical Econometrics*, ed. by B. Baltagi. Basil Blackwell, Oxford.

RIBAR, D. (1992): "Child Care and the Labour Supply of Married Women," *Journal of Human Resources*, 27,??

ROSENZWEIG, M. R., AND K. I. WOLPIN (1980): "Life-Cycle Labor Supply and Fertility: Casual Inferences from Household Models," *Journal of Political Economy*, 88(2), 328–348.

SCHULTZ, P. T. (1978): "The Influence of Fertility on Labor Supply of Married Women: Simultaneous Equation Estimates," in *Research in Labor Economics*, ed. by R. G. Ehrenberg. Jai Press Inc.

WEEKS, M. (1996): "Testing the Binomial and Multinomial Choice Models Using Cox's Non-Nested Test," *Journal of the American Statistical Association (Papers and Proceedings)*.

WEEKS, M., AND C. ORME (1999): "The Statistical Relationship Between Bivariate and Multinomial Choice Models," Department of Applied Economics Working Paper 9912, University of Cambridge.

WHITE, H. (1982): "Regularity Conditions for Cox's Test of Nonnested Hypothesis," *Journal of Econometrics*, 19, 301–18.

WILLIS, R. (1973): "A New Approach to the Economic Theory of Fertility Behavior," *Journal of Political Economy*, 81.

ZELLNER, A., AND T. H. LEE (1965): "Joint Estimation of Relationships Involving Discrete Random Variables," *Econometrica*, 33(2), 382–394.

Figure 1 (a)
Bivariate Model

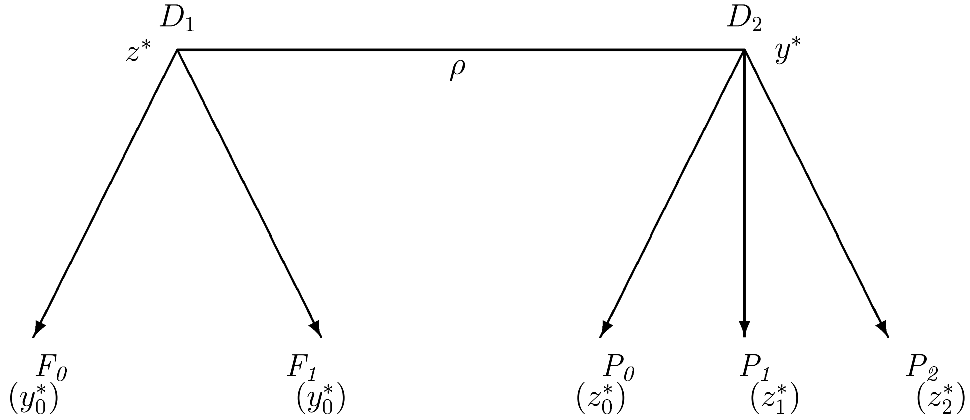

Figure 1 (b)
Multinomial Model

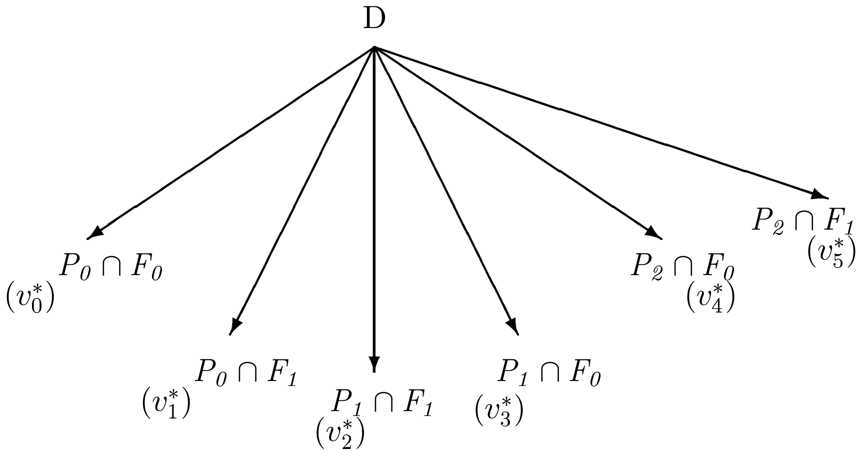

Table 1: Variable Descriptions

Participation =	0 if the woman does not work 1 if the woman works $<$ 30 hours 2 if the woman works $>$ 30 hours
Fertility =	1 if the woman has at least one child between 0 and 15 years old 0 otherwise
Household Income	All receipts from non state pension sources plus all receipts from state benefits plus other transfer (education grants, sickness insurance, maintenance foster allowance) plus income from savings and investment[1]. Plus the husbands usual net annual pay.
Gross hourly wage of women and men	Usual gross pay per month divided by the sum of the number of hours normally worked per week including overtime (adjusted on a yearly base)
Age	Age of the woman in 1994
Age squared	Age squared divided by 100.
Children 0-2	**1** ($0 <$ age of children ≤ 2)
Children 3-4	**1** ($3 <$ age of children ≤ 4)
Children 5-11	**1** ($5 <$ age of children ≤ 11)
Children 12-15	**1** ($12<$ age of children < 15
First degree	1 if first degree or equivalent
A' level	1 if A' level or equivalent; else 0
O' level	1 if O' level; else 0
Working hours	Number of hours normally worked per week
Experience	Sum of months of previous experience in part time, full time and self employed jobs. The part time months were weighted for 50%
Experience squared	Experience squared divided by 1,000
Marital status =	1 if the woman is married, 2 if is cohabiting.

[1] All the receipts refer to the 12 months before the interview (Sep 94).

Table 2: Descriptive Statistics

	Mean	Std Dev.	Min.	Max.	No. of obs.
Participation Decision	0.998	0.830	0	2	2502
Fertility Decision	0.478	0.410	0	1	2502
					2502
Joint Decision					2502
P_0F_0	0.155	0.362	0	1	2502
P_1F_0	0.119	0.324	0	1	2502
P_2F_0	0.248	0.432	0	1	2502
P_0F_1	0.191	0.393	0	1	2502
P_1F_1	0.198	0.394	0	1	2502
P_2F_1	0.096	0.295	0	1	2502
Household annual income (£'s)	11314	9255	0	151820	2502
Marital status	1.165	0.3713	1	2	2502
Age	39.404	10.650	18	60	2502
Age squared	16.661	8.577	3.240	36	2502
Women's predicted gross hourly wage	6.172	1.502	5.48	11.63	2502
Women's observed gross hourly wage	6.139	3.653	0.289	47.402	1615
Women's predicted net hourly wage	5.34	1.11	3.97	17.55	2502
Women's observed net hourly wage	4.785	2.388	0.283	37.471	1615
Predicted working hours	28.497	9.890	1	99	2502
Observed working hours	28.615	11.688	1	99	1615
Children 0-2	0.119	0.324	0	1	2502
Children 3-4	0.096	0.294	0	1	2502
Children 5-11	0.274	0.446	0	1	2502
Children 12-15	0.166	0.372	0	1	2502
First degree	0.094	0.292	0	1	2502
A level	0.182	0.386	0	1	2502
O level	0.284	0.451	0	1	2502
Experience	106.436	99.061	0	542	2502
Experience squared divided by 100	21.138	32.446	0	293.76	2502

Table 3: Wage equation

Dependent variable: female gross hourly wage.

Variables	OLS estimates
Constant	$5.481^{\dagger}$
	(0.263)
First degree	$4.993^{\dagger}$
	(0.282)
A' level	$2.338^{\dagger}$
	(0.231)
O' level	$0.674^{\dagger}$
	(0.198)
Experience	0.000
	(0.007)
Experience squared	0.006
	(0.006)
λ	$-1.376^{\dagger}$
	(0.385)

No. of observations: 1615

R^2 squared 0.199

† denotes coefficient is significant at the 5% level of significance.

λ is the coefficient on the inverse of the Mills Ratio

Table 4: Hours equation

Dependent variable: number of hours usually worked per week

Variables	OLS estimates
Constant	$32.696^\dagger$
	(0.648)
First degree	$4.597^\dagger$
	(0.973)
A level	$3.051^\dagger$
	(0.793)
0 level	$1.993^\dagger$
	(0.684)
Experience	$-0.045^\dagger$
	(0.008)
Experience squared	$0.126^\dagger$
	(0.024)
Children 0-2	$-4.146^\dagger$
	(1.021)
Children 3-4	$-6.215^\dagger$
	1.078
Children 5-11	$-6.433^\dagger$
	(0.670)

No. of observation: 1638
$R^2 = 0.130$
† denotes coefficient is significant at the 5% level of significance.

Table 5: Bivariate Models (2 x 2)

Variables	Compare	Bivariate Logit Estimates†	Bivariate Probit Estimates
Fertility Eqn.	$C/\bar{C}$		
Constant		$-12.811^{\dagger}$	$-7.505^{\dagger}$
		(0.719)	(0.415)
Predicted net wage		-0.017	-0.018
		(0.031)	(0.025)
Household Income		$0.009^{\dagger}$	$0.008^{\dagger}$
		(0.004)	(0.003)
Age		$0.745^{\dagger}$	$0.464^{\dagger}$
		(0.041)	(0.022)
Age squared		$-1.067^{\dagger}$	$-0.660^{\dagger}$
		(0.056)	(0.029)
Participation Equation	W/NW		
Constant		$-1.350^{\dagger}$	$-1.872^{\dagger}$
		(0.652)	(0.396)
Predicted net wage		$0.214^{\dagger}$	$0.210^{\dagger}$
		(0.043)	(0.027)
Household Income		-0.007	-0.005
		(0.005)	(0.003)
Age		$0.086^{\dagger}$	$0.081^{\dagger}$
		(0.031)	(0.019)
Age squared		$-0.144^{\dagger}$	$-0.118^{\dagger}$
		(0.039)	(0.024)
Alfa		$-2.081^{\dagger}$	$-0.380^{\dagger}$
		(0.185)	(0.034)
Log Likelihood		-2818.0	-2813.0

Score test: Unrestricted vs. Restricted MNP model
$P(\chi^2_{(9)} > \chi^2_{calc}) = 0.001$
† denotes coefficient is significant at the 5% level of significance.

Table 6: Bivariate Models (3 x 2)

Variables	Compare	Bivariate Logit Estimates	Bivariate Probit Estimates
Fertility Equation			
Constant	$C/\bar{C}$	$-12.304^{\dagger}$	$-7.615^{\dagger}$
		(0.781)	(0.431)
Predicted net wage		-0.065	$-0.073^{\dagger}$
		(0.042)	(0.025)
Household Income		$0.015^{\dagger}$	$0.010^{\dagger}$
		(0.006)	(0.003)
Age		$0.775^{\dagger}$	$0.486^{\dagger}$
		(0.043)	(0.024)
Age squared		$-1.110^{\dagger}$	$-0.692^{\dagger}$
		(0.058)	(0.033)
Part-time Equation	PT/NW		
Constant		-0.399	-0.032
		(0.598)	(0.681)
Predicted net wage		$0.080^{\dagger}$	$0.109^{\dagger}$
		(0.042)	(0.041)
Household Income		$-0.007^{\dagger}$	-0.002
		(0.004)	(0.004)
Age		$0.071^{\dagger}$	-0.026
		(0.030)	(0.039)
Age squared		$-0.107^{\dagger}$	0.009
		(0.037)	(0.054)
Full Time Equation	FT/NW		
Constant		$-11.262^{\dagger}$	$-4.788^{\dagger}$
		(1.131)	(0.765)
Predicted net wage		$0.869^{\dagger}$	$0.231^{\dagger}$
		(0.004)	(0.040)
Household Income		$-0.017^{\dagger}$	-0.007
		(0.007)	(0.004)
Age		$0.359^{\dagger}$	$0.175^{\dagger}$
		(0.055)	(0.041)
Age squared		-0.455	$-0.210^{\dagger}$
		(0.068)	(0.055)
Alfa		$-0.872^{\dagger}$	$-0.327^{\dagger}$
		(0.039)	(0.018)
Log Likelihood		-3779.4	-4003.1

Table 7: Multinomial Logit: 4 State

Variables	Compare	Estimate
Constant	$W\_\bar{C}/NW\_\bar{C}$	-1.111
		(0.968)
Predicted net wage		$0.196^{\dagger}$
		(0.071)
Household income		-0.011
		(0.008)
Age		$0.158^{\dagger}$
		(0.047)
Age squared		$-0.281^{\dagger}$
		(0.056)
Constant	$NW\_C/NW\_\bar{C}$	$-10.551^{\dagger}$
		(1.367)
Predicted net wage		$-0.224^{\dagger}$
		(0.090)
Household income		0.004
		(0.009)
Age		$0.845^{\dagger}$
		(0.073)
Age squared		$-1.282^{\dagger}$
		(0.096)
Constant	$W\_C/NW\_\bar{C}$	$-17.948^{\dagger}$
		(1.411)
Predicted net wage		$0.309^{\dagger}$
		(0.077)
Household income		0.003
		(0.008)
Age		$1.086^{\dagger}$
		(0.073)
Age squared		$-1.555^{\dagger}$
		(0.095)
Log Likelihood		-2753.8

† denotes coefficient is significant at the 5% level of significant

Table 8: Multinomial Logit 6 State

Variable	Compare	Estimate
Constant	$PT\_\bar{C}/NW\_\bar{C}$	$-6.557^{\dagger}$
		(1.307)
Predicted net wage		$0.436^{\dagger}$
		(0.077)
Household Income		$-0.023^{\dagger}$
		(0.009)
Age		$0.249^{\dagger}$
		(0.061)
Age squared		$-0.312^{\dagger}$
		(0.072)
Constant	$FT\_\bar{C}/NW\_\bar{C}$	-0.975
		(1.036)
Predicted net wage		0.005
		(0.074)
Household income		-0.004
		(0.008)
Age		$0.211^{\dagger}$
		(0.052)
Age squared		$-0.382^{\dagger}$
		(0.064)
Constant	$NW\_C/NW\_\bar{C}$	$-10.451^{\dagger}$
		(1.376)
Predicted net wage		$-0.292^{\dagger}$
		(0.089)
Household income		0.007
		(0.009)
Age		$0.860^{\dagger}$
		(0.073)
Age squared		$-1.304^{\dagger}$
		(0.098)
Constant	$PT\_C/NW\_\bar{C}$	$-19.714^{\dagger}$
		(1.599)
Predicted net wage		$0.378^{\dagger}$
		(0.077)
Household income		-0.001
		(0.009)
Age		$1.147^{\dagger}$
		(0.084)
Age squared		$-1.644^{\dagger}$
		(0.111)
Constant	$FT\_C/NW\_\bar{C}$	$-16.397^{\dagger}$
		(1.852)
Predicted net wage		-0.044
		(0.095)
Household income		0.016
		(0.010)
Age		$1.027^{\dagger}$
		(0.097)
Age squared		$-1.467^{\dagger}$
		(0.127)
Log Likelihood		-3672.6

Table 9: Elasticities of the probability of participation and of having children with respect to the female wage

	Participation	Fertility
Biv. Logit (2×2)		
under 60	0.327	-0.127
under 50	0.223	-0.154
under 40	0.095	-0.217
Biv. Probit (2×2)		
under 60	0.540	0.001
under 50	0.519	-0.024
under 40	0.502	-0.175
Multinomial Logit(4 states)		
under 60	0.569	-0.030
under 50	0.554	-0.054
under 40	0.578	-0.218

Table 10: Elasticities of the probability of participation and of having children with respect to the female wage

	Participation	Fertility	Full-time	Part-time
Biv. Logit (3×2)	0.476	-0.063	-0.293	1.537
Biv. Probit (3×2)	0.583	-0.176	0.444	0.852
Multinomial Logit (6 states)	0.559	-0.040	-0.384	1.595

Table 11: Cox Non-Nested Test Statistics

	H_f: MNL[12]		H_f: Biv-L		H_f: Biv-P	
	H_g: Biv-L	H_g: Biv-P	H_g: MNL	H_g: Biv-P	H_g: Biv-L	H_g: MNL
			I: (2×2) 4 *Choice Models*			
C_{WH}	$-0.677[NA]$	NA	$-11.933[-13.83]$	$-3.924[NA]$	NA	NA
C_{WO}	$-0.566[-4.562]$	-1.133	$-11.768[-9.744]$	$-3.924[-8.018]$	-2.552	-10.814
C_{WX}	$-0.223[-1.308]$	-0.360	$-11.408[-8.570]$	$-3.924[-6.125]$	-2.551	-9.613
			II: (3×2) 6 *Choice Models*			
C_{WH}	-1.430	1.918	-16.418	6.274	NA	NA
C_{WO}	-0.828	1.559	-16.382	5.861	-7.250	-11.394
C_{WX}	-0.491	0.949	-13.823	3.703	-7.183	-11.246