THE UNIVERSITY OF MELBOURNE

DEPARTMENT OF ECONOMICS

RESEARCH PAPER NUMBER 928

MARCH 2005

**DISCRETE HOURS LABOUR
SUPPLY MODELLING: SPECIFICATION,
ESTIMATION AND SIMULTATION**

by

John Creedy
&
Guyonne Kalb

Department of Economics
The University of Melbourne
Melbourne   Victoria   3010
Australia.

Discrete Hours Labour Supply Modelling: Specification, Estimation and Simulation

John Creedy
Department of Economics
The University of Melbourne
E-mail: jcreedy@unimelb.edu.au
and

Guyonne Kalb
Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
Parkville, VIC 3010
Australia
E-mail: g.kalb@unimelb.edu.au

Abstract

The assumption behind discrete hours labour supply modelling is that utility-maximising individuals choose from a relatively small number of hours levels, rather than being able to vary hours worked continuously. Such models are becoming widely used in view of their substantial advantages, compared with a continuous hours approach, when estimating and their role in tax policy microsimulation. This paper provides an introduction to the basic analytics of discrete hours labour supply modelling. Special attention is given to model specification, maximum likelihood estimation and microsimulation of tax reforms. The analysis is at each stage illustrated by the use of numerical examples. At the end, an empirical example of a hypothetical policy change to the social security system is given to illustrate the role of discrete hours microsimulation in the analysis of tax and transfer policy changes.

Keywords: discrete hours labour supply; taxation; microsimulation; maximum likelihood estimation.

1   Introduction

This paper provides an introduction to the analytics of discrete hours labour supply modelling in the context of tax microsimulation modelling. Microsimulation models are used to examine the effects of hypothetical or actual tax and benefit reforms, using a large cross-sectional data set that reflects the degree of heterogeneity found in the population. The level of disaggregation that can be obtained from microsimulation results, compared to what is possible in general equilibrium modelling, is a major advantage when the interest is, for example, in distributional issues such as inequality or poverty or in the effect of reforms on subgroups in the population. Discrete hours models are popular in tax policy microsimulation because it is relatively easy, compared with the continuous hours models, to incorporate taxation and social security details in estimation and simulation.

To reach the stage where a policy change, such as a change in income taxation rates, can be simulated, several steps are needed. First, a model needs to be specified explaining labour supply behaviour. Second, taxation and social security details and individual information on incomes, wages and household composition are needed to calculate net incomes at all possible labour supply levels. Third, the labour supply model is estimated using information on individuals' labour supply, net income at the different labour supply levels and other relevant characteristics. Fourth, once the parameters of the model are estimated, they can be used to simulate the effect of policy changes.[1] In this paper, special attention is given to the three steps of model specification, estimation and microsimulation.

The assumption behind discrete hours labour supply modelling is that utility-maximising individuals choose from a relatively small number of hours

levels, rather than being able to vary hours worked continuously.[2] The discrete approach is perhaps more realistic than the continuous approach, in that typically only a finite number of part-time or full-time working options are available.[3] It is assumed that the same set of hours is available to each individual. The approach also substantially simplifies the information required about the budget set faced by each individual. In the continuous hours context the analysis of choices under piecewise-linear budget lines must deal with the complexities arising from budget sets displaying convex and non-convex ranges, and multiple local equilibria.[4] In practice, the evaluation of the complete range of each individual's unique budget set is cumbersome, given the complexity of most tax and transfer systems. This is further complicated in the case of couples and joint utility maximisation, where the budget constraint is three-dimensional. With discrete hours models it is simply a question of evaluating utility at a small number of points, none of which represents a standard tangency solution.

The advantages of discrete hours modelling are perhaps even stronger in the context of the empirical estimation of individuals' preference functions. With continuous hours modelling several approaches have been adopted.[5] Often a reasonably flexible labour supply function is estimated relating hours worked to net wage rates, non-wage incomes and a range of individual characteristics. The associated utility function is found by appropriate integration methods. Alternatively a supply function is derived from either a direct or (more commonly given the greater flexibility allowed) an indirect utility function. However, considerable problems arise because of the fact that net wages and hours are jointly determined, and problems exist concerning the determination of virtual non-wage incomes for each linear segment of the budget constraint. Indeed, it is extremely difficult to capture the com-

plexities arising from supply behaviour under piece-wise linear constraints in empirical continuous hours models. The discrete hours approach can also deal directly with the population heterogeneity that is a fundamental feature of cross-sectional surveys.

The methods described in this paper are illustrated and reinforced using small numerical examples, for which considerable simplification is obviously required. The first three sections are devoted to specification issues. Section 2 describes the discrete choice modelling framework. In practice, the determinants of any individual's behaviour can never be known with certainty. A feature of the discrete hours approach is that the stochastics are introduced at the initial modelling stage in the utility function rather than in the derived labour supply model; measured utility differs from true utility as a result of measurement, optimisation and other errors. This generates a probability distribution over the set of hours available for work. Section 3 provides a numerical example of the way in which such a probability distribution is generated, where the error terms follow a hypothetical discrete distribution. A more detailed and formal examination of the error specification, and its implications for the probability distribution of an individual's hours worked, is given in section 4. It is worth stressing that tax microsimulations, because of the need to use a large survey data set, are essentially partial equilibrium analyses in which individuals' wage rates are assumed to be fixed. The focus of this type of model is on the supply side of the labour market.

The next three sections concern estimation of the model. Estimation of the parameters of specified preference functions, using the method of maximum likelihood, is considered in section 5. A simplified numerical example of estimation is given in section 6. Alternative specifications of the model are discussed briefly in section 7.

The next two sections are concerned with the use of discrete hours labour supply models in behavioural microsimulation. Section 8 begins by describing the role of calibration, whereby each individual's optimal labour supply in the initial situation corresponds to the discretised hours actually supplied. A simple numerical example of a hypothetical tax reform is then presented. The final subsection discusses the problem of producing summary measures of the income distribution generated by alternative tax systems. A realistic empirical example of a tax reform is presented in section 9, showing how the concepts examined in earlier sections can be applied. Brief conclusions are in section 10.

2 The Basic Model

This section presents the utility maximisation framework and discusses the determination of the probability distribution of hours worked. Subsection 2.1 describes the discrete choice model, involving the introduction of a random term reflecting the difference between actual and measured utility for an individual. In contrast with a deterministic approach, this gives rise to a probability distribution of hours worked for each individual, as discussed in subsection 2.2 and more formally in subsection 2.3. The measurement of labour supply elasticities in this framework is examined in subsection 2.4.

2.1 Utility Maximisation

Consider an individual with a set of measured characteristics, X. The individual (who faces a fixed gross wage rate) maximises utility by selecting the number of hours worked, h, subject to the constraint that only a discrete number of hours levels, h_i, $(i = 1, ..., n)$ are available for work. Utility is determined by the amount of leisure and net income, is increasing in both

arguments, and is bounded by the time and budget constraints. That is, the amount of leisure per week cannot be more than the total amount of time available per week minus the hours of work.[6] Total weekly income is restricted by the available amount of nonlabour and labour income. The latter is the individual's wage rate multiplied by hours worked. Instead of leisure, hours of work are often used as the argument in the utility function because labour supply is typically the key variable of interest in economics.

The utility associated with each hours level is denoted U_i^* and is a function of 'measured' utility $U\left(h_i | X\right)$ plus an error term, v_i, so that:

$$
\begin{aligned}
U_i^* &= U\left(h_i | X\right) + v_i \\
&= U_i + v_i
\end{aligned}
\tag{1}
$$

Although utility is considered to be a function of net income and hours worked, it is not necessary here to refer to net income, since this is determined directly from the associated hours level and the wage and other characteristics of the individual. The term v_i arises from factors such as measurement errors concerning the variables in X, optimisation errors of the individual or the existence of unobserved preference characteristics. Any observation on h is associated with a set of possible draws of the n random variables v_i from their respective distributions. Within this framework, there exists a probability distribution over available hours levels that is influenced by the properties of the v_i. Without these error terms, the model would be deterministic and knowledge of the form of U and the vector X would be sufficient to determined the precise utility-maximising choice of hours level.

The issue considered here is how to generate the probability distribution for labour supply, $p\left(h = h_i\right) = p_i$, for $i = 1, ..., n$, given assumptions about the distributions v_i. In the next sections, emphasis is given to the case where

the errors are independent and identically distributed.

2.2 Probability Distributions

The framework summarised by equation (1) involves a distribution of utility for each discrete hours level, depending on the distributions of the v_i. Suppose for convenience that there are only three hours points. The three probability density functions of U_i^*, for $i = 1, 2, 3$, are shown in Figure 1, where in each case increasing utility involves moving upwards along each vertical axis. The horizontal axis measures relative frequency. The choice of any particular hours level is associated with draws from these three distributions, where the hours level producing the highest U_i^* is chosen.

Consider the probability that hours level h_1 is chosen, given that the value $U_1^* = A$ has been selected from the distribution of U_1^*. This can only be chosen if it is higher than the values of U_2^* and U_3^* selected from their respective distributions. From Figure 1, the probability that $U_2^* < A$ is given by the area B. Similarly the probability that $U_3^* < A$ is the area C. The joint probability that h_1 is chosen, given the selection of $U_1^* = A$, is the probability that $U_2^* < A$ **and** $U_3^* < A$. If the draws from the distributions are independent, this probability is the product, BC, of the two areas.[7]

This relates only to one draw, of $U_1^* = A$, from the distribution of U_1^*. It is necessary to consider the overall probability of h_1 being chosen. This is obtained by adding together all the conditional probabilities, for all possible values of U_1^*.[8] Even for the higher values of U_1^*, Figure 1 suggests that the conditional probabilities of h_1 being selected would in most cases be low. Overall, the probability of h_1 producing maximum utility is small compared to the probability of h_2 or h_3 producing maximum utility.
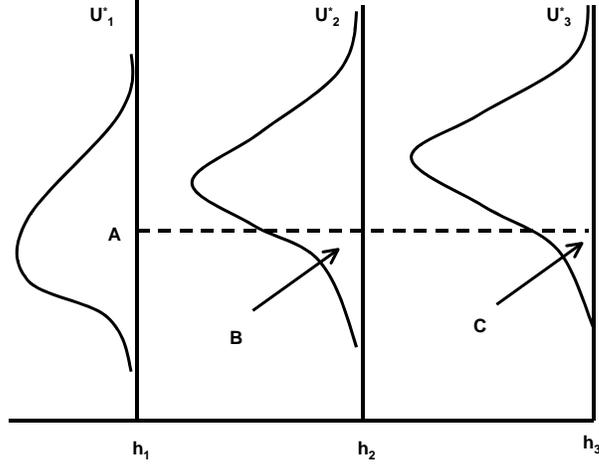
Figure 1: Utility Distributions at Three Hours Levels

2.3 A More Formal Statement

The procedure discussed in the previous subsection is set out more formally here. Consider the hours level, i. Utility maximisation implies that this hours level is chosen if:

$$U_i^* \geq U_j^* \qquad \text{for all } j \tag{2}$$

Substituting for U_i^*, using (1), and rearranging, this condition is equivalent to the requirement that:

$$v_j \leq v_i + U_i - U_j \qquad \text{for all } j \tag{3}$$

Hence, for any given value of v_i, the probability of U_i^* exceeding all other values is equal to the joint probability that $v_i + U_i - U_1 \geq v_1$ and $v_i + U_i - U_2 \geq v_2$ and so on for all j. If the various distributions are independent, this joint probability is the product of the separate probabilities, $P\left(v_j \leq v_i + U_i - U_j\right)$. Therefore, for any given value of v_i, the probability that hours level i produces

9

maximum utility is equal to:

$$\prod_{j \neq i} P\left(v_j \leq v_i + U_i - U_j\right) \tag{4}$$

This is the conditional probability, for a given value of v_i. The overall probability is found by aggregating terms like (4) over all possible values of v_i. The analysis of this problem is considerably simplified by assuming that the form of the distribution of v_i for each i is the same. An example is given in the next section, and this is followed by a more detailed and analytical treatment of the error specification. First, it is necessary to consider the concept of the wage elasticity of labour supply in the discrete hours context.

2.4   Labour Supply Elasticities

The structural basis of the discrete model means that there is no explicit labour supply function which depends directly on wage and other characteristics of the individual.[9] The estimated parameters are those of the utility function, which determine labour supply in terms of a distribution of hours worked. This raises the question of how the concept of the (uncompensated) wage elasticity of labour supply can be applied in the discrete hours context.

An elasticity measure may be based on expected hours worked. Consider an individual with known characteristics, including the hourly wage and the net incomes associated with each hours point, from which the probabilities of being at each of the discrete hours points can be calculated, as shown in more detail in section 4. Using these probabilities the expected value of labour supply can be computed. Next, the individual's gross wage is increased by a small amount, keeping all other characteristics the same, and the new expected labour supply is calculated. An elasticity can be produced by dividing the percentage change in expected labour supply by the imposed

percentage change in the wage. Such elasticities in general vary according to the initial wage rate and the individual's characteristics, as well as the net incomes at the hours points, which are determined by the tax and benefit system.

In some models with more complex error specifications (as discussed in section 7), it is not possible to determine the probabilities analytically. However, a simulation approach can be taken. Values from the relevant error distributions are drawn for all labour supply points, after which the optimal choice of labour supply can be determined by finding the highest U^*. If this process is repeated several times the distribution of labour supply for a particular individual can be determined by counting the number of times each discrete point is the optimal point. Given the probabilities at each of the discrete hours points the expected value of labour supply can be calculated and the process of deriving wage elasticities is then the same as described above.[10]

3   A Numerical Example of Hours Probabilities

This section shows how the probability distribution of an individual's hours of work is derived, using a simple hypothetical numerical example. Suppose for the purposes of this example that v takes only discrete values, v_k, for $k = 1, ..., K$. In general, let $f(a_k)$ denote the proportion of values equal to a_k and let $F(a_k)$ denote the proportion less than or equal to a_k. The value of p_i (the probability of h_i producing the highest utility) is thus obtained as the addition of terms corresponding to (4):

$$p_i = \sum_{k=1}^{K} \left\{ \prod_{j \neq i} F(a_k + U_i - U_j) \right\} f(a_k) \tag{5}$$

Consider a situation in which there are just four levels of labour supply available, so that $n = 4$. The values of U_i associated with each hours level, h_1 to h_4, are respectively 5, 7.5, 10 and 9. For the purpose of this example for a single individual, it is not necessary to specify either the form of the function, U, or the precise discrete hours levels. Clearly, without the stochastic component of utility, h_3 would be chosen.

Suppose that all values for v are drawn independently from the same discrete distribution with four possible outcomes, so $K = 4$, and let v take the values shown in Table 1. In this hypothetical example the arithmetic mean value of v is non-zero.

Table 1: Hypothetical Discrete Distribution of the Error Term

k	1	2	3	4
v	-2.0	0.0	2.0	3.5
$f(v)$	0.1	0.3	0.4	0.2
$F(v)$	0.1	0.4	0.8	1.0

The selection of an hours level is, as explained in section 2, associated in this case with the random draws from the four distributions, each identical to the one shown in Table 1. For example, a set of random values for v_1 to v_4 may be say 2, 0, -2 and 2 respectively. These give rise to utilities, U_i^*, of 7, 7.5, 8, and 11 for the hours levels h_1 to h_4 respectively. Hence h_4 is chosen in this case. Given a draw of -2 from the distribution of v_3, the option h_4 can dominate h_3 (that is, $U_4^* > U_3^*$) if v_4 takes either of the values 0, 2 or 3.5. The conditional probability of h_4 dominating, given this selection from v_3, is thus $0.3 + 0.4 + 0.2 = 0.9$, found by adding the relevant values of $f(v)$ in Table 1. The enumeration of all possible combinations of this type is most efficiently carried out following the approach underlying equation (5).

Consider the probability of selecting hours level h_3. The relevant values

are shown in Table 2. The second column, headed $U_3 - U_j$ shows the differences in the values of U; these are all positive, as hours level 3 has, by assumption, the highest value of U. The column headed $k = 1$ relates to the values and probabilities when a_1 is drawn for v_3. The first row shows that when $k = 1$, that is when $v_3 = a_1 = -2$, the term v_1 must be less than $a_1 + U_3 - U_1 = 3$ in order to ensure that hours level 3 has a higher value of U^*. From the assumed distribution in Table 1, there is a probability of 0.8 that v is less than 3. This is shown in the second row of Table 2. Similarly, when $k = 1$, hours level h_3 gives higher utility than h_4 only if $v_4 < -1$; this has a probability of 0.1.

<p align="center">Table 2: Conditional Probabilities For Hours Level 3</p>

j	$U_3 - U_j$		$k = 1$	$k = 2$	$k = 3$	$k = 4$
1	5	$a_k + U_3 - U_1$	3	5	7	8.5
		$F(a_k + U_3 - U_1)$	0.8	1	1	1
2	2.5	$a_k + U_3 - U_2$	0.5	2.5	4.5	6
		$F(a_k + U_3 - U_2)$	0.4	0.8	1	1
4	1	$a_k + U_3 - U_4$	-1	1	3	4.5
		$F(a_k + U_3 - U_4)$	0.1	0.4	0.8	1
Conditional probability that						
$U_3^* > U_1^*$ and $U_3^* > U_2^*$ and $U_3^* > U_4^*$			0.032	0.320	0.800	1.0

The conditional probability that h_3 is chosen, when $v_3 = -2$, is therefore $(0.8)(0.4)(0.1) = 0.032$. The final column of Table 2 shows that when $k = 4$, so that $v_3 = a_4 = 3.5$, hours level 3 always dominates and the conditional probability of it being chosen is 1. The overall probability p_3 is thus given

by:

$$
\begin{aligned}
p_3 &= \sum_{\substack{k=1 \\ k \neq 3}}^{4} P(U_3^* > U_j^*, \text{for all } j \neq 3 \mid v_3 = a_k) \, p(v_3 = a_k) \\
&= \sum_{\substack{k=1 \\ k \neq 3}}^{4} \left(\prod_{j \neq 3} F(a_k + U_3 - U_j) \right) f(a_k) \\
&= (0.032)\,(0.1) + (0.32)\,(0.3) + (0.80)\,(0.4) + (1.0)\,(0.2) \\
&= 0.619
\end{aligned}
\tag{6}
$$

Similar calculations show that $p_1 = 0.003$, $p_2 = 0.080$ and $p_4 = 0.298$. The resulting probability distribution of hours clearly depends in a complex way on the distribution of the error term.

This example has been constructed in order to illustrate the way in which the hours distribution for an individual is derived from the underlying stochastic specification and utility levels. In practice more structure has to be imposed by specifying a precise form for the error distribution, $f(v)$. A special case using a continuous distribution is examined in the next section, which is necessarily more technical than the previous discussion.

4 Specification of the Error Distribution

This section derives the probability distribution of hours worked for a special case of the distribution of error terms. This special distribution results in a multinomial logit model for utility. This model has been used extensively in discrete choice modelling. The discrete error distribution in the previous section was used only for convenience, and it is first necessary to restate the problem where v is considered to be a continuous random variable. Hence, $f(v)$ and $F(v)$ are now the density and distribution functions respectively of v. It is possible to convert the result in equation (5) into the following

form for continuous v, remembering that hours continue to be discrete:

$$p_i = \int_{-\infty}^{+\infty} \left\{ \prod_{j \neq i} F\left(v_i + U_i - U_j\right) \right\} f\left(v_i\right) dv_i \qquad (7)$$

The expression in (7) takes all the possible conditional probabilities, represented by $\prod_{j \neq i} F\left(v_i + U_i - U_j\right)$, and integrates v_i out to obtain the required marginal distribution p_i. Given that the conditional probabilities require the product of distribution functions, $F\left(.\right)$, it cannot be expected that an arbitrary choice of $f\left(v\right)$ would be tractable. This section considers a special case generating a highly convenient form for the hours distribution.

4.1 A Special Case: The Extreme Value Distribution

Suppose the distribution of v is described by the following density function:

$$
\begin{aligned}
f\left(v\right) &= e^{-v} e^{-e^{-v}} \\
&= \exp\left(-v - e^{-v}\right) \qquad (8)
\end{aligned}
$$

for which the distribution function is:

$$F\left(v\right) = e^{-e^{-v}} \qquad (9)$$

The choice of this thin-tailed distribution has the obvious advantage that no further parameters need to be estimated.[11] This is known as an Extreme (Maximum) Value Type I distribution, often abbreviated simply to 'extreme value' distribution. This is because it has been found useful in many applications involving extreme values.[12] It is highly tractable in the present context. These qualities have generally been taken as sufficient justification for its use, though section 7 briefly discusses some alternatives.

The arithmetic mean of this distribution is non-zero, being equal to 0.5772 (Euler's number); the mode is zero and the median is $-\ln\left(\ln 2\right)$. The shape

of the distribution is illustrated in Figures 2 and 3, showing the density and distribution functions respectively. In Figure 3, the distribution function used in the numerical example of section 3 is shown for comparison: this is obviously a step function for the discrete distribution.
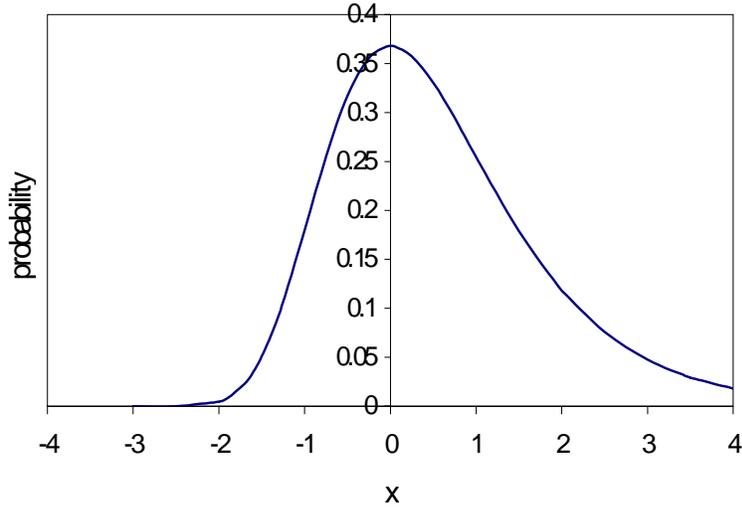

Figure 2: Extreme Value Probability Density Function

Substitution into $\left[\prod_{j \neq i} F\left(v_i + U_i - U_j\right) \right] f\left(v_i\right)$ gives:

$$\exp\left(-v_i - e^{-v_i}\right) \prod_{j \neq i} \exp\left(-e^{-v_i} e^{-U_i + U_j}\right) \tag{10}$$

Noting that the logarithm of $\prod_{j \neq i} \exp\left(-e^{-v_i} e^{-U_i + U_j}\right)$ can be expressed as $-e^{-v_i} \sum_{j \neq i} e^{-U_i + U_j} = -e^{-v_i} e^{-U_i} \sum_{j \neq i} e^{U_j}$, and using $x = \exp\left(\log x\right)$, the expression in (10) becomes:

$$\exp\left[-v_i - e^{-v_i}\left(1 + e^{-U_i} \sum_{j \neq i} e^{U_j}\right)\right] \tag{11}$$

Furthermore, say

$$1 + e^{-U_i} \sum_{j \neq i} e^{U_j} = e^{U_i} e^{-U_i} + e^{-U_i} \sum_{j \neq i} e^{U_j} = e^{-U_i} \sum_{j=1}^{n} e^{U_j} = e^{\lambda_i}.$$
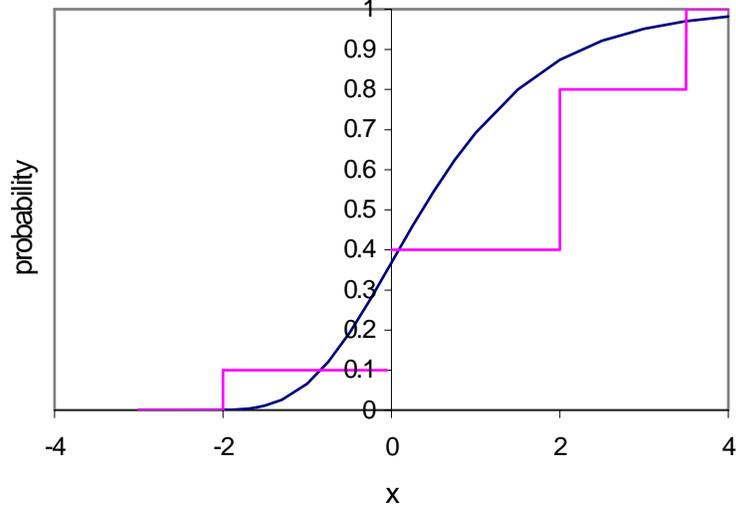
16

Figure 3: Extreme Value Cumulative Distribution Function and the Discrete Distribution from Table 1

Hence (11) can be rewritten more succinctly as $\exp\left[-v_i - e^{-(v_i-\lambda_i)}\right]$. Thus:

$$p_i = \int_{-\infty}^{+\infty} \exp\left[-v_i - e^{-(v_i-\lambda_i)}\right] dv_i \tag{12}$$

Further simplification is achieved using the transformation, $v_i' = v_i - \lambda_i$, so that $v_i = v_i' + \lambda_i$ and $dv_i = dv_i'$, whereby:

$$
\begin{aligned}
p_i &= \int_{-\infty}^{+\infty} \exp\left[-v_i' - \lambda_i - e^{-v_i'}\right] dv_i' \\
&= e^{-\lambda_i} \int_{-\infty}^{+\infty} \exp\left[-v_i' - e^{-v_i'}\right] dv_i' \\
&= e^{-\lambda_i} \int_{-\infty}^{+\infty} f\left(v_i'\right) dv_i' \\
&= e^{-\lambda_i} \\
&= \frac{e^{U_i}}{\sum_{j=1}^{n} e^{U_j}} \tag{13}
\end{aligned}
$$

In this special case, the probability distribution of hours of work for an individual depends in a convenient way on the measured utility levels asso-

17

ciated with each hours level.[13] In the deterministic framework, monotonic transformations of the utility function have no effect on the choice of optimal hours worked. However, in the present context such transformations (other than the addition of a constant) affect the probabilities associated with each hours level. The discrete choice model flowing from the assumption of an extreme value distribution is called a multinomial logit model. More on the logit model and alternative discrete choice models can be found in Maddala (1983).

5 Parameter Estimation

The previous sections have examined the discrete choice model underlying an individual's labour supply behaviour. The basic assumption in the labour supply model is that individuals maximize utility which depends on two arguments, income and hours of work. Utility is expected to increase with income and to decrease with hours of work (or increase with the complement of working hours, leisure time).

This section discusses how this model can be estimated with the help of data, using the maximum likelihood method. An advantage of the discrete hours framework, in contrast to the continuous approach, is that it can be applied to any legitimate utility function. Hence, no explicit assumption about utility functions is made in the present section: their specification is discussed in section 7. The extreme value error distribution, examined in the previous section, is used. The construction of the likelihood function is described in subsection 5.1 and its maximisation is considered in subsection 5.2.

5.1 The Likelihood Function

The notation used in the previous sections did not need to distinguish between individuals, since only a single individual was examined. However, estimation uses information from a cross-section of individuals. Suppose there are M individuals and the index k is used to refer to individuals $k = 1, ..., M$. There are, as before, n discrete hours levels h_i, for $i = 1, ..., n$. It is first necessary to indicate the optimal hours level for the kth person; denote this by h_{i_k}, so that i_k indicates the chosen value of i (the hours index) for person k. The probability of selecting this hours level is p_{i_k} and the corresponding optimal utility level is $U^*_{i_k,k}$. All other utility levels (associated with other hours levels) are denoted $U^*_{j,k}$ for $j = 1, ..., n$.

Using this notation:

$$p_{i_k} = P\left(U^*_{i_k,k} \geq U^*_{j,k} \text{ for all } j\right) \tag{14}$$

Thus when all v_i are assumed to follow the extreme value distribution discussed in Section 4.1, the probability associated with the optimal hours chosen by person k is expressed as:

$$p_{i_k} = \frac{e^{U_{i_k,k}}}{\sum_{j=1}^{n} e^{U_{j,k}}} \tag{15}$$

The joint probability that individual 1 selects h_{i_1} **and** individual 2 selects h_{i_2} **and** individual 3 selects h_{i_3} and so on, is given, assuming that the decisions are made independently, by the product:

$$P(h_{i_1}, ..., h_{i_M}) = p_{i_1} p_{i_2} ... p_{i_M} = \prod_{k=1}^{M} \frac{e^{U_{i_k,k}}}{\sum_{j=1}^{n} e^{U_{j,k}}} \tag{16}$$

This joint probability concerns the probability of the set of hours levels, h_{i_k} for $k = 1, ..., M$, being chosen by the M individuals, given their preferences

and other personal characteristics, and assuming that all v follow identical extreme value distributions.

Researchers do not know the parameters of the assumed form of preference functions, but have information about the hours worked by each individual in a random sample taken from the population. In addition, data are available on personal characteristics and net incomes of each individual at each discrete hours point. The net incomes are not observed directly but are obtained from knowledge of each individual's wage rate and the details of the tax and transfer system. The taxation and benefit rules are applied to the gross income of each individual at each of the discrete points to obtain the associated net income. Depending on the complexity of the rules and the data available, it may not be possible to include all benefits. Furthermore, the wage rates of those who are not in employment at the time of the survey cannot be observed, so it is necessary to impute wage rates for non-workers using estimated wage functions.[14]

The probability in (16) can be viewed from another perspective. Given an assumption about the general form of the utility functions, it is possible to find parameter values that, if true, would produce the highest probability of observing the actual hours values. The expression in (16) is reinterpreted as being a function of the unknown parameter values, for a given set of observed hours. Since the framework is one in which a particular true set of parameters is assumed to exist, and any variations are attributed to sampling variations, it is not appropriate, when discussing the function in terms of parameters, to refer to a probability of parameters taking particular values. Rather, it is correct to refer to the probability of observing this particular sample of individuals (with their combinations of characteristics and hours worked) conditional on the parameter values. Suppose that each individual's

utility function depends on a vector of coefficients β, with elements β_ℓ, for $\ell = 1, ..., S$. The probability statement in (16) can be rewritten as:

$$L(\beta_1, ..., \beta_S) = \prod_{k=1}^{M} \frac{e^{U_{i_k,k}}}{\sum_{j=1}^{n} e^{U_{j,k}}} \tag{17}$$

where $L(\beta_1, ..., \beta_S)$, a function of the unknown parameters (for a given sample of individuals for whom hours worked are observed), is referred to as the Likelihood Function. Here the fixed parameters are effectively treated as if they were variables. The estimates, $\widehat{\beta}_1, ..., \widehat{\beta}_S$, produced by finding values for $\beta_1, ..., \beta_S$ that maximise the value of this function, are referred to as maximum likelihood estimates.

Taking logarithms gives the log-likelihood for this model:

$$\log L = \sum_{k=1}^{M} \left[U_{i_k,k} - \log \left(\sum_{j=1}^{n} e^{U_{j,k}} \right) \right] \tag{18}$$

This monotonic transformation does not affect the maximum likelihood estimates but, by converting products into sums, makes analysis easier.

5.2   Maximum Likelihood Estimation

The log-likelihood is maximised when the following first-order conditions are satisfied:[15]

$$\frac{\partial \log L}{\partial \beta_\ell} = 0 \text{ for all } \ell = 1, ..., S \tag{19}$$

Differentiation of (18) gives:[16]

$$\begin{aligned} \frac{\partial \log L}{\partial \beta_\ell} &= \sum_{k=1}^{M} \left[\frac{\partial U_{i_k,k}}{\partial \beta_\ell} - \left(\frac{1}{\sum_{j=1}^{n} e^{U_{j,k}}} \right) \sum_{j=1}^{n} \frac{\partial e^{U_{j,k}}}{\partial \beta_\ell} \right] \\ &= \sum_{k=1}^{M} \left[\frac{\partial U_{i_k,k}}{\partial \beta_\ell} - \sum_{j=1}^{n} \left(\frac{e^{U_{j,k}}}{\sum_{j=1}^{n} e^{U_{j,k}}} \right) \frac{\partial U_{j,k}}{\partial \beta_\ell} \right] \end{aligned} \tag{20}$$

In considering the terms in (20) it should be remembered that, even if every individual has the same general form of utility function, the individual utilities depend on the personal characteristics in X.

Using (19) and (20), to rewrite the first-order conditions, gives for all ℓ:

$$\sum_{k=1}^{M} \left[\frac{\partial U_{i_k,k}}{\partial \beta_\ell} - \sum_{j=1}^{n} p_{j,k} \frac{\partial U_{j,k}}{\partial \beta_\ell} \right] = 0 \tag{21}$$

This has the interpretation that the aim of this method is to make the first derivatives of utility in the observed hours points on average equal to the weighted average of derivatives of utility over all possible hours points. The weights for each individual are equal to the probabilities of each discrete hours level. Although this is an interesting interpretation of the first-order conditions, it does not provide practical help in trying to solve the nonlinear set of equations.

The solution, giving the set of maximum likelihood estimates for all β_ℓs, can be obtained using numerical methods involving a sequence of iterations which lead efficiently from an arbitrary starting point to the solution. In particular, Newton's method involves repeatedly solving the following matrix equation, where $\beta^{[I]}$ denotes the vector of parameters in the Ith iteration:

$$\beta^{[I+1]} = \beta^{[I]} + \left[-\frac{\partial^2 \log L}{\partial \beta_\ell \partial \beta_s} \right]^{-1}_{\beta^{[I]}} \left[\frac{\partial \log L}{\partial \beta_\ell} \right]_{\beta^{[I]}} \tag{22}$$

and the first and second derivatives are evaluated using the parameters $\beta^{[I]}$. Furthermore, it can be shown that the inverse of the matrix of second derivatives at the final iteration provides an estimate of the variance-covariance matrix of parameter estimates. Further discussion of Newton's method, which is often used to solve nonlinear equations, can be found in the appendix.

The application of Newton's method in the present context requires the second derivatives of the likelihood function. Differentiating (21) again with

respect to parameter β_s, gives:

$$\frac{\partial^2 \log L}{\partial \beta_\ell \partial \beta_s} = \sum_{k=1}^{M} \left[\frac{\partial^2 U_{i_k,k}}{\partial \beta_\ell \partial \beta_s} - \sum_{j=1}^{n} \left\{ p_{j,k} \frac{\partial^2 U_{j,k}}{\partial \beta_\ell \partial \beta_s} + \frac{\partial U_{j,k}}{\partial \beta_\ell} \frac{\partial p_{j,k}}{\partial \beta_s} \right\} \right] \qquad (23)$$

where:

$$\frac{\partial p_{j,k}}{\partial \beta_s} = p_{j,k} \left[\frac{\partial U_{j,k}}{\partial \beta_s} - \sum_{t=1}^{n} p_{t,k} \frac{\partial U_{t,k}}{\partial \beta_s} \right] \qquad (24)$$

An example using this procedure is described in the following section.

6   A Numerical Example of Estimation

This section illustrates the application of the maximum likelihood method using a simple numerical example involving a linear form of utility function. Although this example is obviously unrealistic as it has few individuals and a simple utility specification, the very simplicity helps to concentrate on the mechanics of the method: the general approach is no different in a more realistic empirical context. Utility is assumed to be independent of an individual's characteristics except for hours worked, wage and other income. Appropriate allowance for dependence on characteristics is discussed in section 7. Here, all individuals have the same utility function with the same parameters. This does not mean that all individuals are expected to have the same optimal level of hours. Firstly, people with different wage levels have different levels of income at the hours points and so optimal hours are located at different points. Secondly, the error term v introduces random differences in utility caused by unobserved factors.

Suppose the utility functions take the simple linear form:

$$U^* = \beta_h h + \beta_y y + v \qquad (25)$$

This implies that the marginal utility of net income is constant and equal to β_y and the marginal utility of hours worked is constant and equal to β_h.

The latter coefficient is expected to be negative. For each individual, the chosen hours point is observed and the value of y for each discrete labour supply point h can be calculated, given information on the gross wage of each individual and knowledge of the tax and transfer system.

Suppose also that there are only three individuals, whose details are shown in the top section of Table 3. There are just three hours levels available for work, 0, 20 and 40 hours, corresponding to not working at all, working part time and working full time respectively. The observed gross wage rates, conditional on the observed hours of work for each individual, give the gross income shown in the final column. In this example, the individual with the highest wage rate works longer hours.[17]

Table 3: Three Individuals and Three Hours Levels

Person	Gross wage	Chosen hours	Gross income
1	4	0	0
2	8	20	160
3	10	40	400
Incomes at alternative hours			
		Hours	Income
1	4	20	80
	4	40	160
2	8	0	0
	8	40	320
3	10	0	0
	10	20	200

Assuming for simplicity that there are no income taxes or benefit payments, the net income is simply equal to gross income. Estimation requires information about the net incomes at all the possible hours levels for each individual. In the present example the required values are easily obtained, as shown in the lower section of Table 3. Where there is a complex tax and

transfer system, the required net incomes are obtained by applying the rules of the system, along with an assumption regarding the take-up of benefits, at each hours level.

The simplified form of utility function in (25) means that the required derivatives of U with respect to parameters are easy to evaluate, since $\partial U/\partial \beta_h = h$ and $\partial U/\partial \beta_y = y$. Appropriate substitution of these values, along with values from Table 3, into the first-order conditions in (21) gives:[18]

$$
\begin{aligned}
\left[\frac{\partial \ln L}{\partial \beta_h}\right]_{\widehat{\beta}} &= 0 - \frac{20\exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 80) + 40\exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 160)}{1 + \exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 80) + \exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 160)} \\
&\quad +20 - \frac{20\exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 160) + 40\exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 320)}{1 + \exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 160) + \exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 320)} \\
&\quad +40 - \frac{20\exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 200) + 40\exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 400)}{1 + \exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 200) + \exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 400)} \\
&= 0 \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (26)
\end{aligned}
$$

$$
\begin{aligned}
\left[\frac{\partial \ln L}{\partial \beta_y}\right]_{\widehat{\beta}} &= 0 - \frac{80\exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 80) + 160\exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 160)}{1 + \exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 80) + \exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 160)} \\
&\quad +160 - \frac{160\exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 160) + 320\exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 320)}{1 + \exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 160) + \exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 320)} \\
&\quad +400 - \frac{200\exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 200) + 400\exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 400)}{1 + \exp(\widehat{\beta}_h 20 + \widehat{\beta}_y 200) + \exp(\widehat{\beta}_h 40 + \widehat{\beta}_y 400)} \\
&= 0 \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (27)
\end{aligned}
$$

The result is two nonlinear equations in the two unknowns $\widehat{\beta}_y$ and $\widehat{\beta}_h$. Using an iterative solution procedure, as described in section 5.2, the maximum likelihood estimates were found to be $\widehat{\beta}_h = -15.41$ and $\widehat{\beta}_y = 1.93$.[19]

Consider the wage elasticity of labour supply for person 2, defined as in subsection 2.4 in terms of changes in expected hours. At the observed

wage level, the hours and corresponding net incomes (equal to gross incomes since by assumption there are no taxes in this example) in Table 3 are used, with the parameter estimates, to obtain the utilities corresponding to each hours point, by appropriate substitution in $U_{i2} = \widehat{\beta}_h h_i + \widehat{\beta}_y y_{i2}$. From these, the probabilities of being at each of the labour supply points are given by $p_{i2} = \exp(U_{i2}) / \sum_j \exp(U_{j2})$ and expected labour supply $E(h(2))$ is calculated using $E(h(2)) = \sum_i p_{i2} h_i$. In this example $E(h(2)) = 0.16 \times 0 + 0.30 \times 20 + 0.54 \times 40 = 27.6$.

After increasing the wage by 1 per cent, new net incomes and hence new utilities for each discrete hours point are obtained. Using the resulting new probabilities, expected hours are found to be 39.5. This implies an elasticity of 43. In this simplified example, with just three individuals, two positive hours levels and linear utility, a small change in the wage rate for this individual leads to a very high probability of working for 40 hours. In practice, given the substantial heterogeneity in populations, considerable differences in individual elasticities are typically found.[20]

7 Alternative Specifications

This section presents a number of alternative specifications of the basic model discussed above. The discussion in this section is meant as an overview only and provides much less detail than the discussion in the previous sections. However, the basic approach to estimation would remain similar for several of these alternative specifications.

First, the form of utility functions is examined in subsection 7.1. Allowance for participation in welfare programmes, often described in terms of the 'take-up' of benefits, is examined in subsection 7.2. Alternative ways in which allowance may be made for individuals' personal characteristics are

discussed in subsection 7.3. The effect of the nature of a particular discrete hours point is described in subsection 7.4. Finally, alternatives to the use of the extreme value distribution are briefly discussed in subsection 7.5.

7.1   Utility Functions

It has been mentioned that the discrete hours approach offers considerable flexibility in the form of utility function that can be used. The linear form used in the numerical example of estimation is obviously highly restrictive. The assumption of constant marginal utilities is implausible and in empirical applications, utility functions usually allow for diminishing marginal utility. A popular choice in applied work is the quadratic utility function:

$$U = \beta_y y + \beta_h h + \alpha_y y^2 + \alpha_h h^2 + \alpha_{yh} yh \tag{28}$$

where the marginal utility of income is:

$$\frac{\partial U}{\partial y} = \beta_y + 2\alpha_y y + \alpha_{yh} h \tag{29}$$

For example, Keane and Moffitt (1998) have used this specification. An alternative is the translog specification in which the arguments of utility are income and leisure $(l = T - h)$, rather than income and hours worked:

$$U = \beta_y \ln y + \beta_l \ln(T-h) + \alpha_y (\ln y)^2 + \alpha_l (\ln(T-h))^2 + \alpha_{yl} \ln y \ln(T-h) \tag{30}$$

where the marginal utility of income is:

$$\frac{\partial U}{\partial y} = \frac{\beta_y + 2\alpha_y \ln y + \alpha_{yl} \ln(T-h)}{y} \tag{31}$$

This specification has been used by, for example, Van Soest (1995).

Both specifications allow for diminishing returns through the quadratic terms. Thus, if α_y is negative the marginal utility of income decreases with

the amount of income. Furthermore the cross-product term allows for complementarity between income and leisure, when α_{yh} is negative or α_{yl} is positive, or substitutability, when α_{yh} is positive or α_{yl} is negative. For example, the value of income may decrease if less leisure time is available, that is extra income may be appreciated less if there is no time for consumption.

Neither the translog nor the quadratic utility function is automatically quasi-concave across the full range of possible parameter values. This is not a problem as long as the optimal parameter values result in a utility function that is quasi-concave in the observed labour supply points. This contrasts with continuous hours labour supply modelling, where the necessary restriction of the parameter space may bias substitution effects upwards and income effects downwards, and which in addition is cumbersome in maximum likelihood estimation; as explained in MaCurdy, Green and Paarsch (1990). In discrete hours labour supply modelling, it is sufficient to check for quasi-concavity after estimation, which involves two necessary conditions, discussed by Van Soest (1995).[21]

The quadratic and translog utility functions can both be easily extended to allow for households consisting of couples, where both partners simultaneously determine labour supply. This is achieved by assuming that the couple maximises one utility function, which seems a reasonable assumption for households where the members pool their incomes. However, a common criticism of this type of model is that the assumption of one common utility function for the household as a whole is not realistic. Unfortunately, alternatives using bargaining models and other types of non-unitary collective models require detailed data and their own set of assumptions which are needed, for example, to divide consumption into shared and private goods or to construct a sharing rule for income. This approach has been used in, for

example, Chiappori (1988), Bourguignon and Chiappori (1994), Browning et al. (1994), Apps and Rees (1997) and Blundell et al. (1998). Such models need to be simplified in other areas. As a result, researchers who focus on tax and benefit policy issues and are interested in incorporating the full detail of tax and benefit systems have mostly chosen unitary utility functions.

The quadratic utility function for a couple can be written as:

$$U = \beta_y y + \beta_m h_m + \beta_f h_f + \alpha_y y^2 + \alpha_m h_m^2 + \alpha_f h_f^2 + \alpha_{ym} y h_m + \alpha_{yf} y h_f + \alpha_{fm} h_f h_m$$

(32)

where the index m denotes hours and parameters of the male and the index f denotes hours and parameters of the female, and y represents joint income. The parameter α_{fm} indicates whether the male's and female's labour supply are complements or substitutes

7.2   Welfare Participation

The utility function can be extended through addition of a term for welfare participation, or benefit take-up, following Moffitt (1983). The choice between discrete labour supply points is then extended to a choice between discrete labour supply points with and without welfare participation, whenever relevant. It is expected that disutility is attached to participation in welfare programmes, which could be caused for example by the costs of applying for benefits. These could be pecuniary costs or non-pecuniary costs, such as the time needed to travel to a social security office, or by a psychological effect, where people on welfare feel stigmatised. The latter explanation is more likely to be important when participation in welfare is clearly noticeable to the outside world, such as through payment in shops with Food Stamps in the U.S.

A simple way of adding welfare participation to the utility function is through the addition of a dummy variable for participation. Examples of this approach can be found in Fraker and Moffitt (1988), Hoynes (1996), Hagstrom (1996), Smith (1997), Keane and Moffitt (1998) and Kalb (1999, 2000). Let $d_w = 1$ if the person participates and $d_w = 0$ if the individual does not take-up the benefit, even if entitled to it. The coefficient on this variable indicates the disutility associated with participation in welfare; that is, a larger negative value indicates greater disutility. For the quadratic utility function, the specification is therefore:

$$U = \beta_y y + \beta_h h + \alpha_y y^2 + \alpha_h h^2 + \alpha_{yh} yh + \phi d_w \qquad (33)$$

The participation parameter can be made dependent on individual characteristics in the same way as for the preferences for work or income. This is described in the following subsection.

An alternative approach is to estimate an unordered model of moving from one choice to another, where the amount of labour supply and participation in the welfare programme jointly determine choice. In this specification there is no explicit welfare participation parameter, but the gain in utility from a choice with welfare participation compared with a choice without welfare participation can be determined conditional on the income gain associated with the move between these choices and other individual characteristics. Bingley and Walker (1997, 2001) estimated a three-point labour supply model where at all, some or none of the labour supply points there is the additional option of participation in a welfare programme.

7.3 Personal Characteristics

Consider again the linear utility function:

$$U^* = \beta_h h + \beta_y y + v \tag{34}$$

This can be extended to make the preference parameters dependent on personal and household characteristics, such as education, number and age of children or an individual's own age, which are likely to influence the preference for work and income. Including these characteristics in the preference for work parameter, the utility function could be written as follows:

$$U^* = \left(\beta_{h1} + \beta_{h2} age + \beta_{h3} d\right) h + \beta_y y + v \tag{35}$$

where, say, $d = 1$ if the age of the youngest child is 0 to 4, and $d = 0$ otherwise. In this case, two extra parameters for the preference for work are included, so the likelihood now depends on four unknown parameters which need to be estimated.

This specification is more flexible than in the numerical example of section 6, where just one preference parameter for work was estimated. For example, individuals with young children are allowed to have different preferences compared with individuals without young children. This approach can be used to estimate the effect of an individual's characteristics on preferences and may help to explain differences in behaviour between individuals with similar wages but different personal characteristics.

This addition means that the effects on wage elasticities of labour supply (as defined above in terms of expected hours worked) of characteristics like age or household composition can be examined. Expected labour supply can be calculated for two individuals who are the same except for the characteristic of interest. There is thus scope for a wide range of elasticities, which provide information on the effect of specific characteristics.

The approach reflected in equation (35) does not incorporate unobserved heterogeneity of individuals because allowance is made only for the measured characteristics. This can be overcome by adding unobserved heterogeneity to the preference parameters. Hence the coefficient on h is written as:

$$\beta_h = \beta_{h1} + \beta_{h2}age + \beta_{h3}d + \varepsilon \tag{36}$$

This introduces an additional error term, ε, which is typically assumed to be normally distributed. This addition complicates the method of estimation somewhat, so that the method of simulated maximum likelihood is now required. However, estimation of such models, including correlated error terms in the different preference terms, remains fairly straightforward using this method; as is explained, for example, in Van Soest (1995).

7.4 Characteristics of Hours Points

It is often observed that the probability of obtaining a job offer depends on the desired number of hours of work. Euwals (2001) showed that there is a discrepancy between observed and desired hours of work, which converge only to some extent over time. This indicates that some individuals work a suboptimal number of hours, which is however preferred over not working. For example, finding a job of 5 hours per week may be more difficult than finding a 40-hour job. As a result some discrete hours points are not well-represented by the standard labour supply model, which does not allow for demand side restrictions. For example, it is often found that labour supply models overpredict part-time hours of work. Several methods have been used to overcome this lack of fit to the observed labour supply. Some examples of alternative approaches are discussed briefly here.

First, an *ad hoc* approach of including a penalty parameter for particular hours of work in the utility function has been used to reduce the utility at

certain hours points, so that the probability at these hours points is reduced; examples include Van Soest (1995), Callan and Van Soest (1996) and Kalb (2000). A second approach involves the inclusion of the probability of a job offer at the different discrete hours points in the model, which can be applied when desired hours of work are known. For example, Woittiez (1991) used the hours restrictions as a way of specifying a discrete model, that is only a subset of discrete points have positive probability of being in the choice set of the individual. The possible choices in this subset can be determined from the actual hours, because all other available job offers must have had a lower utility. Alternatively, Euwals and Van Soest (1999) took desired labour supply as given and examined the probability of obtaining job offers at the different hours points separately.

Third, a parameter measuring the fixed cost of working can be subtracted from net income in a quadratic utility function; see for example, Duncan and Harris (2002). This approach is similar to the first approach but is expressed in dollars rather than units of utility. Thus it is intuitively more appealing, although the costs represented by this parameter are both pecuniary and non-pecuniary costs.

In the fourth approach, the number of job offers in an interval associated with the discrete point is directly used to weight the probabilities derived by using the extreme value distribution. See Aaberge, Dagsvik and Strøm (1995), Aaberge, Colombino and Strøm (1999) and Kornstad and Thoresen (2002, 2003), who use this approach. A final approach, by Duncan and Harris (2002a), adapts the multinomial logit model to allow for captivity at particular discrete hours points. This increases the probability of observing an individual at particular hours points. It allows some hours points to have a high probability, which does not need to depend on an individual's

33

characteristics. This may, for example, be expected at the standard full-time 40-hours point.

7.5   Alternative Error Distributions

The use of the extreme value distribution contains an assumption that has, in previous sections, remained implicit. This form assumes that there is no correlation between the error terms of the different hours alternatives. This is usually referred to as the 'independence of irrelevant alternatives' property, and means that taking out one of the choices would not affect the odds ratios of the other choices. For example, suppose that individuals can initially choose between 0, 5, 10, 15, ..., 45 and 50 hours of work. Taking out the 10 hours choice, it seems unlikely that the relative probabilities of the other choices would not change. If 10 hours were no longer an option, it seems likely that individuals previously preferring this discrete point would move to the neighbouring labour supply points, thus changing the odds between the choices. An obvious but unpopular approach is to extend the extreme value distribution with fixed mean and variance to a version where these two parameters are estimated, allowing for correlation between the choices.

Some authors have chosen an alternative to the extreme value distribution for the random term to be added to measured utilities. This usually complicates estimation and a sign of the larger complexity is that in such cases it has been possible to distinguish at most three discrete hours points. This contrasts with around ten hours points for each individual, when using the extreme value distribution. This remains possible even when labour supply is estimated jointly for couples. However, the advantage of the alternative approaches is that greater flexibility is allowed in modelling the relationship between the labour supply of two members of a couple or between labour

34

supply and welfare participation.

A model related to the multinomial logit described in the previous sections is the nested multinomial logit. Hagstrom (1996) showed that this specification allows for correlation between some of the decisions in the model. In his application, correlation between the wife's labour supply choice and welfare participation within the husband's labour supply choice is allowed. This relaxes the independence assumption between all alternatives in the standard multinomial logit model, although some structure is still imposed on the covariance matrix. In addition, a distinction is made between choice-specific variables and individual-specific variables, imposing more structure on the way characteristics influence the different choices by individuals.

An alternative to the extreme value distribution is a normal distribution, which would lead to a probit-type model instead of the logit-type model. However, multivariate probit models are difficult to estimate, even for as few as three categories. An additional problem is that it is impossible analytically to determine the limits of integration which indicate which discrete hours point is preferred. With the recent development of simulation techniques combined with more powerful computers, this type of model has become more feasible and some researchers have explored this option. Fraker and Moffitt (1988) estimate labour supply and participation in two welfare programs for female heads of household in a reduced form model. Three levels of labour supply are distinguished. The choice for these levels of income depends on the preference parameter for work, which depends on an individual's characteristics, and an unobserved factor which is assumed to be normally distributed. No error terms are added directly to the utility function. The model can be estimated because the ranges for the preference parameter where each hours point is optimal can be written down.[22] This

only works when the budget constraint is not too nonconvex, which might otherwise make it impossible for part-time work to be optimal in this specification of the model. The problem of finding the limits of integration, which determine which discrete labour supply point and whether welfare participation is chosen, necessitated the reduced form approach by Fraker and Moffitt. A similar specification using a structural approach can be found in Keane and Moffitt (1998), who overcome the problems with the limits of integration by using advanced simulation techniques. With the simulation approach there is no need to determine analytically the limits of integration. However, estimation is cumbersome and time consuming.

Bingley *et al.* (1995) use an approach where the difference between utility levels is modelled rather that the utility function itself. Under the assumption of normally distributed error terms on the utility function, a multinomial probit model can be derived. They distinguish three discrete points and model the probability of preferring non-participation over part-time employment and the probability of preferring non-participation over full-time employment. Thus the distribution of the differences in utility between non-participation and part-time employment and between non-participation and full-time employment are modelled. They allow for correlation across the choices. The variance-covariance matrix is normalized by assuming that the variance of the difference between the part-time and full-time error term has a variance of one. When more than three choices are specified, simulation techniques are needed for the estimation.

Finally, a flexible non-parametric approach was taken by Hoynes (1996) who added unobserved heterogeneity to the preference parameters for labour supply of husband and wife and for welfare participation. This approach uses a discrete factor representation, where sets of M different pairs of unobserved

heterogeneity for the husband's and wife's preferences for work parameter and for the preference for welfare participation $(\theta_{hk}, \theta_{wk}, \mu_k)$ are observed with a probability π_k where $k = 1, ..., n$ and $\sum_{k=1}^{M} \pi_k = 1$. The flexibility of this approach is appealing, but it adds a large number of additional parameters to be estimated ($4M - 1$ in addition to the number of parameters in a multinomial specification). For large M, any correlation between the different error terms can be represented by this specification. In addition to this discrete probability distribution which is meant to capture the correlation between the different preference terms, normally distributed independent error terms are added to the preference for welfare participation and the observed hours of work. The use of error terms for the hours is an interesting approach to circumvent the need to group observed hours in categories with more or less arbitrary boundaries. Inputs in Hoynes's model are continuous hours, and the difference between these continuous observed hours and the discrete labour supply points is accounted for through a multiplicative factor $\exp(\varepsilon)$, where ε is normally distributed with mean $-\sigma^2/2$ and variance σ^2. Hence zero hours are observed with certainty, but positive hours are observed with an error. Although the intuition behind this model is simple, estimation of the model is difficult, particularly for large M.

8   Tax Reforms and Simulation

The previous sections have all concentrated on the specification and estimation of the discrete hours labour supply model. This section turns to the use of such models in behavioural tax microsimulation. Policy changes for which this can be done are mostly of a financial type, such as a change in the amount of benefits, the withdrawal rate, eligibility for benefits, or the range of income where a withdrawal rate applies. Such changes result in a

change in net income at each of the discrete hours points, which may result in a shift in the optimal choice for an individual. These changes contrast with, for example, modifications to rules regarding the duration of benefits, residence requirement, willingness to accept training, the ability to refuse job offers, and reasons for job loss. These are important design features of a social security system, but are difficult to accommodate in microsimulation.

First, subsection 8.1 describes the method of calibration used to place individuals in their (pre-reform) observed discretised hours level under the tax system in operation at the time of the survey. The generation of a post-reform probability distribution of hours worked for each individual, conditional on them being at their observed pre-reform hours, is also described. Secondly, subsection 8.2 provides a small numerical example of a tax reform, using the three hypothetical individuals used in the illustration of maximum likelihood estimation. Thirdly, subsection 8.3 turns to the question of how to obtain inequality and poverty measures when a distribution of hours worked, rather than a specific level, is available for each individual.

8.1 Individual Calibration

Once the parameters of the specified preference functions have been estimated, they can be used to simulate the effects on labour supply of policy changes. A common approach is to use a base data set and start from the labour supply observed in this data set to obtain a starting point for simulation based on the observed labour supply under a particular tax and benefit system. This is achieved by calibration, which means that error terms are drawn from the relevant distribution (for example, the extreme value distribution) and added to the measured utility in each of the hours points. If this results in the observed labour supply being the optimal choice for the indi-

vidual, the draw is accepted; otherwise another set of error terms is drawn and checked. This is repeated until the required number of sets of error terms are drawn.

These sets of error terms, that are possible values leading to the observed labour supply, are then used to compute a distribution of labour supply after a specified reform. The more error terms that are drawn, the more accurate is the computed distribution, especially for those points with low probability. Given the individual's characteristics and draws for the error term, utility at each hours level after the change can be determined. In this way, a probability of being in each of the discrete hours points, conditional on the pre-reform labour supply, can be derived for each individual.

8.2   A Numerical Example

This section presents a small tax policy simulation using the example from subsection 6, in order to illustrate the procedure described above. The utility for all individuals is the estimated utility function $U = -15.41h + 1.93y$, obtained for the simple illustrative example of section 6. In the simulation, a linear benefit and tax system is introduced. Individuals without income receive 15 units of income and gross income (excluding this basic income of 15) is taxed at 20 per cent. This is sometimes described as a basic income - flat tax structure, a social dividend scheme, or a negative income tax. Table 4 presents income and measured utility at the three discrete hours points for all three individuals before and after the reform.

>From the table it is clear that the introduction of the tax system has made work much less attractive. Adding random draws from the extreme value distribution to the estimated utility function, in order to obtain the U^*s, results in different utility levels for each draw. Here only the draws

Table 4: Utility Pre- And Post-Reform

h	y_1	U_1	y_2	U_2	y_3	U_3
			Pre-reform			
0	0	0	0	0	0	0
20	80	-153.8	160	0.6	200	77.8
40	160	-307.6	320	1.2	400	155.6
			Post-reform			
0	15	28.9	15	28.9	15	28.9
20	79	-155.7	143	-32.2	175	29.5
40	143	-340.4	271	-93.4	335	30.2

are kept that result in the observed hours being the optimal choice for the relevant individual. Table 5 presents, for each individual, ten sets of draws from the extreme value distribution which result in the observed hours being the optimal choice for each individual. The corresponding utility levels are presented below each value of v, where U^{*0} indicates utility pre-reform.

Calculation of the utility conditional on these draws, after the reform has been introduced, results in the post-reform utility levels, indicated by U^{*1}. From the measured utility levels, U, in Table 4, individuals 1 and 2 are most likely not to participate whereas individual 3 has utility levels at 0, 20 and 40 hours of work which are relatively close to each other. In Table 5 it can be seen that in draw 9 the post-reform utility of individual 3 is highest for 20 hours of work and in draw 4 it is highest at zero hours of work, whereas in the other draws the utility is highest when the person is working full time. For the other two individuals, non-participation always results in the highest utility.

The results from these ten draws can be summarised in a transition table. Table 6 presents such a matrix for this example. The last column presents the distribution of labour supply before the reform and the last row presents this distribution after the reform. The distribution before the reform consists

Table 5: Utility Pre-Reform and Post-Reform for Ten Sets of Accepted Draws From the Extreme Value Distribution

	Person 1			Person 2			Person 3		
	$h =0$	20	40	0	20	40	0	20	40
$v_1^{(a)}$	-1.07	1.36	0.18	3.00	3.49	0.23	1.17	1.03	2.43
U^{*0}	-1.07	-152.44	-307.42	3.00	4.09	1.43	1.17	78.83	158.03
U^{*1}	27.88	-154.37	-340.23	31.95	-28.72	-93.15	30.12	30.58	32.58
v_2	-0.80	0.44	-0.01	0.78	0.24	-1.42	0.91	-0.78	5.68
U^{*0}	-0.80	-153.36	-307.61	0.78	0.84	-0.22	0.91	77.02	161.28
U^{*1}	28.15	-155.29	-340.42	29.73	-31.97	-94.77	29.86	28.77	35.83
v_3	0.23	-0.74	0.034	-0.17	1.28	0.08	0.80	1.23	0.99
U^{*0}	0.23	-154.54	-307.56	-0.17	1.88	1.28	0.80	79.03	156.59
U^{*1}	29.18	-156.47	-340.37	28.78	-30.92	-93.29	29.75	30.78	31.14
v_4	2.55	2.40	-0.02	-0.64	1.63	0.52	2.07	1.25	0.46
U^{*0}	2.55	-151.40	-307.62	-0.64	2.23	1.72	2.07	79.05	156.06
U^{*1}	31.50	-153.33	-340.43	28.31	-30.57	-92.85	31.02	30.80	30.61
v_5	-0.02	0.66	1.26	0.71	3.74	2.41	-0.71	-0.40	-0.33
U^{*0}	-0.02	-153.14	-306.34	0.71	4.34	3.61	-0.71	77.40	155.27
U^{*1}	28.93	-155.07	-339.15	29.66	-28.47	-90.96	28.23	29.15	29.82
v_6	0.06	1.63	-1.27	0.11	2.43	-0.41	-1.24	-0.67	-0.53
U^{*0}	0.06	-152.17	-308.87	0.11	3.03	0.79	-1.24	77.13	155.06
U^{*1}	29.01	-154.10	-341.68	29.06	-29.78	-93.78	27.71	28.88	29.61
v_7	-0.62	1.08	-0.55	1.20	0.84	-1.50	1.77	1.12	2.31
U^{*0}	-0.63	-152.72	-308.15	1.206	1.44	-0.30	1.77	79.32	157.91
U^{*1}	28.32	-154.65	-340.96	30.15	-31.37	-94.87	30.72	31.07	32.46
v_8	0.14	1.23	0.17	-0.51	1.85	1.04	-1.35	-0.55	1.12
U^{*0}	0.14	-152.57	-307.43	-0.51	2.45	2.24	-1.35	77.24	156.72
U^{*1}	29.09	-154.50	-340.24	28.44	-30.35	-92.33	27.60	28.99	31.27
v_9	2.75	-0.53	0.36	0.16	1.04	-0.22	0.63	0.43	-0.69
U^{*0}	2.75	-154.33	-307.24	0.16	1.64	0.98	0.63	78.23	154.90
U^{*1}	31.69	-156.26	-340.05	29.11	-31.17	-93.59	29.58	29.98	29.45
v_{10}	1.73	-0.33	-1.19	-1.24	1.48	-0.86	-0.50	0.19	0.23
U^{*0}	1.73	-154.13	-308.79	-1.24	2.08	0.34	-0.50	77.99	155.83
U^{*1}	30.68	-156.06	-341.60	27.71	-30.73	-94.23	28.45	29.74	30.38

note a: v_i indicates the error term for draw i, which is added to the calculated
utility level before and after the reform

41

of the percentages of individuals observed in each of the hours points. The distribution after the reform is constructed from the individual probabilities of being at each of the discrete hours points. After the reform an individual cannot be assigned to one of the discrete hours points, but has a positive probability of being at each of the hours points. However, some of these probabilities may be close to zero. All these probabilities for an individual add up to one. The numbers inside the matrix are row percentages indicating the probability of individuals moving from one discrete hours point to another. Thus, the probability of moving from zero hours is nil, the probability of moving from 20 hours to zero hours is 100 per cent and the probability of remaining at 40 hours is 80 per cent. There is a 10 per cent probability of moving from 40 to zero hours and the probability of reducing labour supply from 40 to 20 hours is also 10 per cent.

Table 6: Labour Supply Transition Matrix

| | Hours post-reform | | | |
Hours pre-reform	0	20	40	Distribution
0	100	0	0	33.333
20	100	0	0	33.333
40	10	10	80	33.333
Distribution	70.000	3.333	26.667	100

Instead of simulating the probabilities, the unconditional probability can also be calculated. The predicted unconditional probability of person 3 being at zero hours, 20 hours and 40 hours is 15.4 per cent, 28.1 per cent and 56.5 per cent respectively. These probabilities are calculated by computing

$$P(h_1) = \frac{\exp(U_{1,1})}{\exp(U_{1,1}) + \exp(U_{2,1}) + \exp(U_{3,1})} = \frac{\exp(28.9)}{\exp(28.9) + \exp(29.5) + \exp(30.2)} = 0.154, \text{ and}$$

similar expressions for the other hours points. These are unconditional probabilities, but given the large difference between utility at the different hours levels in the starting situation and the observed hours being the optimal

42

hours, there should not be much difference between the conditional and unconditional probabilities in this case, because most draws from the extreme value distribution would be accepted. The simulation method using draws from the extreme value distribution provides results that are different from these expected probabilities. Table 6 shows that these were 10, 10 and 80 per cent respectively for 0, 20 and 40 hours of work. However, by increasing the number of draws in the simulation method the approximation becomes more accurate. For example for 1000 draws, the percentages at 0, 20 and 40 hours is 15.8, 29.5 and 54.7 per cent.

Using a similar simulation approach, wage elasticities can be calculated for the three individuals in the example. These can be computed with and without calibration. Table 7 show the results of using the alternative methods for each individual in the pre-reform situation. At the wage levels of persons 1 and 3 a small change does not have any effect on the relative utility levels at each of the hours points. Therefore no change in labour supply is expected. However, for person 2 the utility levels of the three hours points are closer to each other. As a result, a small change in the wage level has a large effect on expected labour supply. It is only for person 2 that calibration has an effect on the outcomes. For the other two persons, nearly all possible draws of the error term result in the observed labour supply choice, whereas for person 2 the error term can shift the optimal outcome from one point to another. Here it is shown that calibration can make a difference to the result. Using calibration in this example, the expected wage elasticity is about twice as large as without calibration. This example produces large elasticities because only two positive hours levels are used and the utility levels at those hours levels are close together.

Table 7: Expected Hours and Wage Elasticities of Labour Supply: Simulated Approach

	Person 1	Person 2	Person 3
Wage rate	4	8	10
	Calibrated results		
Expected hours at original wage	0	20	40
Expected hours after 1% wage increase	0	38.42	40
Wage elasticity of labour supply	0	92.1	0
	Non-calibrated results		
Expected hours at original wage	0	27.72	40
Expected hours after 1% wage increase	0	39.66	40
Wage elasticity of labour supply	0	43.1	0

8.3 Income Distribution Measures

As shown in the previous subsection, the discrete hours microsimulation approach does not identify a particular level of hours worked for each individual after a policy change, but generates a probability distribution over the discrete hours levels used. As a result, the usual formulae for poverty and inequality measures cannot be applied directly. Suppose there are M individuals and n possible outcomes of labour supply (hours levels) for each individual. This results in n^M possible combinations of labour supply, and thus income distributions, each of which results in a different value for poverty and inequality measures.

Under the reasonable assumption that individuals' distributions of hours are independent, the probability of each income distribution, or outcome P_q, is given by the product of the relevant probabilities. Hence, if $p_{i,j}$ is the probability that individual j is at hours level i, the joint probability P_q is equal to $p_{i,1}p_{j,2}...p_{r,M}$, where q runs from 1 to n^M; and i, j, r can attain values between 1 and n (indicating the labour supply points chosen in combination q for each individual). In principle, each inequality or poverty

measure can be calculated as a weighted average of the measures over all possible outcomes, with weights equal to the probabilities P_q. However, for any realistic sample size, even for few discrete labour supply points, the large number of possible combinations makes it computationally impractical to calculate all n^M distributions and associated probabilities P_q.

Instead of examining all combinations of income levels, it would be possible to adopt a sampling approach. A large number of possible income distributions could be obtained by taking random draws from each individual's hours distribution. With a sufficiently large number of randomly selected samples, the proportion of each hours combination would replicate the precise probabilities discussed above. But this approach still requires a large computational effort, depending on the number of draws needed to obtain a good approximation. In considering alternatives which offer more practical solutions, the most obvious is perhaps a simple approach where the expected income is calculated for each individual and is used as if it were a single representative level of income for that individual; see, for example, Gerfin and Leu (2003).

A further approach, proposed by Creedy, Kalb and Scutella (2003), treats all possible outcomes for every individual as if they were separate observations. The outcomes are weighted by the individual probabilities of labour supply to produce a pseudo distribution. Consider the pseudo income distribution with Mn income levels, each associated with a corresponding probability. The incomes are $y_{h,i}$, where h ranges from 1 to n and i ranges from 1 to M, and associated probabilities are $p_{h,i}/M$. The division by M ensures that the sum of the probabilities adds to 1. The $y_{h,i}$ values are placed in a single vector, $z = \{y_{h,i}\}$ with Mn elements, with the associated probabilities

45

given by $p' = \{p_{h,i}/M\}$. Hence:

$$\sum_{j=1}^{Mn} p'_j = \sum_{i=1}^{M} \sum_{h=1}^{n} p_{h,i}/M = 1 \tag{37}$$

Simulation experiments have shown that this pseudo distribution approach performs better than the two alternatives discussed above.

9 An Empirical Example Using Australian Data

Previous sections have concentrated on the theory of labour supply specification, estimation and microsimulation, reinforced by simplified numerical examples. This section presents a realistic example of those microsimulation methods. The results presented here concentrate on labour supply effects, but of course it is also possible to evaluate the implications for government taxes and expenditures or to examine inequality and poverty effects.[23]

The simulation results were obtained using the Melbourne Institute Tax and Transfer Simulator (MITTS), and this is described very briefly in subsection 9.1. Subsection 9.2 describes the tax change examined and gives examples of summary results.

9.1 The MITTS Model

MITTS is a microsimulation model based on the Australian Survey of Income and Housing Costs (SIHC). The SIHC is a representative sample of the Australian population, containing detailed information on labour supply and income from different sources, in addition to a variety of background characteristics of individuals and households. For present purposes the 1998 tax system was used as the starting point because the most recent SIHC available at the time of analysis was held in 1997/1998. The econometric estimates of

preferences underlying the behavioural responses were estimated using data from this period as well. Pre-reform net incomes at alternative hours levels are based on the MITTS calculation of entitlements, not the actual receipt. Hence it is assumed that the take-up rates are 100 per cent.[24] All results are weighted using the weights provided with SIHC to represent the population level effects.

The labour supply responses generated by MITTS-B, the behavioural component of MITTS, are based on quadratic preference functions where the parameters are allowed to vary with individuals' characteristics. These parameters were estimated for five demographic groups, which include married or partnered men and women, single men and women, and sole parents. For couples, labour supply is estimated simultaneously for the two members. Different sets of discrete hours points are used for men and women. Given that the female hours distribution is much more spread over part-time and full-time hours than the male distribution, which is mostly divided between non-participation and full-time work, women's labour supply is divided into eleven discrete points, whereas men's labour supply is represented by three points. For individuals in the data set who are not working, an imputed wage is obtained, based on estimated wage functions, which allow for possible selectivity bias.

As explained in the previous section, the behavioural simulations begin by recording the discrete hours level for each individual that is closest to the observed hours level. Then, given the parameter estimates of the quadratic preference function, a random draw is taken from the distribution of the error term, in this case the extreme value type I distribution. This draw is rejected if it results in an optimal hours level that differs from the discretised value observed. The accepted drawings are used in the determination of the

optimal hours level after the policy change. A total of 100 accepted draws are produced, giving a probability distribution over the set of discrete hours for each individual under the new tax structure.[25]

9.2   Labour Supply Changes

The Australian tax and transfer system has a large number of means-tested benefits. The hypothetical policy change analysed here reduces the benefit taper or abatement rates in the 1998 tax structure to 30 per cent. All taper rates of 50 per cent and 70 per cent are reduced to 30 per cent, while leaving all basic benefit levels unchanged.[26] A 30 per cent taper rate means that for every dollar of additional income in the household, the benefit payment is reduced by 30 cents.

The effect on labour supply of this reduction in the taper rates is equivocal because it does not automatically mean a reduction in effective marginal tax rates for all individuals. This is an inevitable consequence of flattening the marginal rate structure while keeping basic benefit levels unchanged. A summary of the labour supply effects is given in Table 8 for the five demographic groups.[27] After the reform, more sole parents are expected to participate in the labour market since very few women move from work to non-participation, whereas a substantial proportion moves into work from non-participation. The net effect is more than 8 per cent. However, there is a relatively small negative effect for a subgroup caused by the 1.8 per cent of sole parents who decrease their labour supply, which is partly counteracted by the 1.3 per cent of sole parents who increase their working hours after the reform. Nevertheless, the resulting average weekly hours are still increased by nearly 3 hours showing that the overall effect is positive.

The changes in the probability of working, for several categories, are pre-

Table 8: Summary of Labour Supply Responses

Behavioural Response	Couples:		Single	Single	Sole
	Men	Women	Men	Women	Parents
Workers[a] (per cent, base)	56.93	40.63	52.29	43.56	41.28
Workers[a] (per cent, reform)	57.49	40.24	52.55	43.44	49.54
non-work–>work (per cent)	1.29	1.72	0.33	0.07	8.3
work–>non-work (per cent)	0.73	2.10	0.06	0.18	0.03
Workers working more	0.08	0.37	1.07	0.42	1.29
Workers working less	2.60	1.43	0.01	0.44	1.81
Average hours change	-0.37	-0.49	0.32	-0.10	2.88

Note a: self employed workers are not included in the workers group

sented in Table 9. Since actual labour supply points are rounded to the nearest discrete labour supply point, the definition of non-participation depends on the number of discrete labour supply points that are used. For women it is working less than 3 hours and for men it is working less than 10 hours. Sole parents are predicted to have a larger increase in the probability of working as a result of reduced taper rates than other groups. This sensitivity to work incentives is found in several other studies (Blundell et al., 2000; Blundell and Hoynes, 2000). Families with more children seem also more likely to participate in the labour market after the reform.

An example of a transition matrix, produced following the method described in the previous section, is shown in Table 10. This is for sole parents, who can be seen to experience both increases and decreases in labour supply as a result of the change. Compared with singles and couples (not shown here), sole parents are more likely to change labour supply, particularly at the lower and upper end of the hours range. Sole parents working fewer than 25 hours seem most likely to increase their hours whereas sole parents working 35 hours or over are more likely to reduce their hours.

Table 9: Changes in Work Probability

	Percentage point change						
	decrease of			increase of			
	(10-50)	(2-10)	0	2-10	10-50	>50	Mean
Income unit type							
Couple	2	18	72	5	3	-	-0.37
Couple+dep.child	-	6	19	55	12	8	0.60
Single	0	3	96	1	0	0	0.09
Sole parent	0	0	46	12	41	-	8.27
Number of children							
None	1	11	84	3	1	0	-0.14
One	4	17	59	9	12	-	1.46
Two	4	14	64	10	8	-	0.85
Three	4	10	67	10	8	-	1.05
Four	4	6	65	12	12	-	2.15
Five	-	3	64	8	25	-	5.48
Six	-	2	79	9	9	-	1.54

Table 10: Labour Supply Transition Matrix: Sole Parents (row percentages)

Pre-	Post-reform hours											Total
reform	0	5	10	15	20	25	30	35	40	45	50	
0	85.9	0.0	0.3	0.8	1.0	1.9	1.8	1.7	2.1	2.2	2.5	58.7
5	-	89.2	0.3	0.4	1.0	0.7	1.7	1.9	1.6	2.3	0.8	2.9
10	-	-	86.3	0.6	0.5	1.4	1.5	2.7	2.1	2.4	2.5	2.2
15	-	0.1	0.1	91.7	0.3	0.7	1.4	1.1	1.7	1.4	1.6	2.8
20	-	-	-	0.1	95	0.2	1.1	0.9	0.8	1.3	0.7	3.9
25	0.0	-	-	-	-	97.8	0.0	0.3	0.4	0.6	0.9	3.6
30	-	0.4	0.3	0.3	0.5	1.1	95.2	0.2	0.7	0.8	0.4	3.4
35	-	0.2	0.3	0.7	1.0	1.6	0.5	94.7	0.2	0.6	0.3	4.0
40	0.2	0.3	0.6	1.6	1.7	2.5	1.7	0.3	90.8	0.2	0.2	12.3
45	0.4	0.3	0.4	1.0	1.9	1.8	2.0	1.9	1.1	87.9	1.3	1.0
50	0.1	0.3	0.4	0.8	1.1	1.0	1.1	1.3	0.9	0.4	92.5	5.2
Total	50.5	2.7	2.2	3.3	4.6	5.2	4.7	5.1	12.7	2.5	6.5	100

10 Conclusions

This paper has provided an introduction to the basic analytics of discrete hours labour supply modelling to be used in microsimulation. Special attention was given to model specification, estimation and microsimulation. The field is of course subject to continual development. For example, several extensions are occurring with regard to the specification of the different random error terms in the utility function, which are aimed at increasing the flexibility of the labour supply model. Alternative models relax the assumption of particular restrictive patterns in the variance-covariance matrices of the error terms in use, such as independence between the different labour supply choices. An increase in computing power has made some of these extensions feasible, although they are often still quite burdensome to carry out.

One area related to the discussion in this paper, that has received little attention in the literature so far, is concerned with the evaluation of simulation outcomes. One of the strengths of tax microsimulation modelling is that it is based on a sample representing the heterogeneity of the population. This allows researchers to analyse the impact of changes on the full range of individuals present in the population of interest. This full range of different individuals allows the measurement of the effect of changes on for example inequality of poverty at a level of detail which would be impossible with methods that use more aggregated data. When using discrete choice labour supply models in simulation, the outcomes of analyses are probabilistic in nature. Measures of welfare dealing with these probabilistic outcomes need further development.

This paper has given several numerical examples to illustrate the more technical exposition of the methods used in this research field. Finally, an

empirical example was given to illustrate the kind of information that can be provided in order to enhance policy debates. In microsimulation it is easy to calculate effects for subgroups, because all effects are calculated at the individual level before being aggregated into the population level or alternatively into subgroups of the population with particular characteristics. This provides a large degree of flexibility in the way the results can be presented. The microsimulation approach also allows a wide variety of effects to be calculated, as long as the base sample has sufficient information for the calculation. Here the effect on labour supply was shown. However, other effects such as the effect on government expenditure on benefit payments or government revenue from income taxation, or the effect on inequality in the population could be calculated using this approach. The microsimulation method also provides a convenient way to analyse the effect of separate components of complex policy changes. It is suggested that the approach offers much potential for further interesting and valuable applications and extensions.

Notes

[1]Creedy *et al.* (2002) discuss microsimulation modelling in detail. Examples of microsimulation studies are Bingley *et al.* (1995), Scholz (1996), Blundell *et al.* (2000), Bingley and Walker (2001), Duncan and Harris (2002), Creedy, Kalb and Kew (2003), Gerfin and Leu (2003).

[2]Early influential papers on discrete choice modelling include McFadden (1973, 1974) and it seems that the first to use a discrete approach to labour supply modelling were Zabalza, Pissarides and Barton (1980).

[3]Van Soest, Woittiez and Kapteyn (1990) and Tummers and Woittiez (1991) show that a discrete specification of labour supply can improve the representation of actual labour supply compared to a continuous specification.

[4]Simulation requires either a search over all segments and corners of each individual's constraint, or the use of an algorithm such as that described by Creedy and Duncan (2002).

[5]The first generation of labour supply models linearised the budget constraint by taking the average net wage rate or the marginal wage rate in the observed hours. This results in a simple regression model if an appropriate utility function is chosen. This type of model is of limited use when interest is in policy analysis related to the tax and benefit system. The second generation of models examined the full budget constraint when searching for optimal labour supply, allowing for any nonlinearities and nonconvexities. Burtless and Hausman (1978) were the first to use this approach; see Hausman (1979, 1985) or Moffitt (1986) for a discussion of the approach.

[6]Most models implicitly allow for home production by assuming that leisure includes home production time. Few articles explicitly allow for home production given the measurement problems. Exceptions are, for example, Becker (1965), Wales and Woodland (1977), Kooreman and Kapteyn (1987), Apps (1994), and Apps and Rees (1996, 1997).

[7]The standard rule for independent probabilities is that $P(A \text{ and } B) = P(A) P(B)$.

[8]The appropriate combination of probabilities here follows the general rule that $P(A \text{ or } B) = P(A) + P(B)$.

[9]This contrasts with the continuous hours approach where a supply function arises from utility maximisation subject to the budget constraint. However, the need to be able to move between utility and labour supply functions in continuous hours microsimulation places a severe restriction on the range of functions used.

[10]A variation on this approach, in which the individual's observed hours at the actual wage are initially used rather than the expected hours, is discussed in section 8 below.

[11]If, instead of the additive form used here, the multiplicative form $U_i^* = U_i v_i$ were adopted, with $f(v) = \frac{1}{v} \exp\left(-\frac{1}{v}\right)$, equivalent results would be obtained, as can be seen by taking logarithms and transforming the distribution.

[12]If a process that generates values for a variable is observed over a period and the maximum value observed is set equal to x, the resulting distribution of x can often be described by the above form. The more general form is $f(x) = \frac{1}{\beta} \exp\left(-\frac{x-\mu}{\beta}\right) \exp\left(-e^{-\frac{x-\mu}{\beta}}\right)$. The standardised form therefore has $\mu = 0$ and $\beta = 1$. This distribution is also referred to as a Gumbel, or double exponential, or Fisher-Tippett Type I distribution. There is a corresponding extreme minimum value distribution.

[13]For the numerical example considered earlier, the hypothetical measured utility levels for the four hours points are 5, 7.5, 10 and 9. Substitution into (13) gives the probabilities 0.005, 0.056, 0.686, and 0.253.

[14]An alternative is to estimate a joint wage and labour supply model (see for example,

Gerfin, 1993).

[15]The second-order sufficient conditions for the solution to represent a maximum are not examined here.

[16]Remembering that $\partial \log x / \partial x = 1/x$, and using the function of a function rule.

[17]If allowance were made for other characteristics, unobserved heterogeneity, and measurement or optimisation errors, an economic model could be derived where this does not always need to be the case. Some low-wage individuals may prefer long hours, and some high-wage individuals may prefer shorter hours, depending on their characteristics.

[18]This specification for v automatically takes care of the scaling of utility, because $\frac{\exp(U_1)}{\exp(U_1)+\exp(U_2)} \neq \frac{\exp(aU_1)}{\exp(aU_1)+\exp(aU_2)}$. Therefore no normalisation is needed when using this approach.

[19]The iterative process was started from a value of 0.01 for both parameters. The only prerequisite for starting values is that the function is defined for those values. When dealing with exponentials, as in this example, large starting values are not recommended because of potential overflow problems. No standard deviations are calculated given that the example consists of three individuals only; the matrix of second derivatives is poorly-conditioned.

[20]Studies based on continuous hours models typically report only a single elasticity evaluated at the sample averages of relevant variables, so care should be taken in making comparisons.

[21]The first is the basic requirement that utility increases with income. The second condition is more complicated but straightforward to check.

[22]The calculation of these boundaries is based on two indifference curves. The first obtains bounds such that $U(0, y_0) = U(20, y_{20})$ and the second imposes $U(20, y_{20}) = U(40, y_{40})$.

[23]In examining revenue implications, the average values over the successful calibration draws for each individual are used. To analyse distributional implications, the technique described in section 8.3, using the successful draws only, is needed.

[24]Furthermore, MITTS applies only income tests, as there is at present no asset imputation in the model. However, income from assets is included in the income tests.

[25]In some cases, 100 successful random draws producing observed hours as the optimal hours cannot be generated within a reasonable number of total draws. If after 5000 draws, the model fails to predict the observed labour supply 100 times, the individual is dropped from the simulation and fixed at the observed hours of work. This occurs 521 times, which in addition to the 121 rejected cases because of unrealistic wages represents 6.5 per cent of all individuals in the database

[26]The exception is the withdrawal rate on parental income for people receiving Youth Allowance or AUSTUDY, which remains at 25 per cent.

[27]For further details on the effects and results of this policy reform, see Creedy, Kalb and Kew (2003).

References

Aaberge, R., Dagsvik, J.K. and Strøm, S. (1995) Labor supply responses and welfare effects of tax reforms. *Scandinavian Journal of Economics*, 97, 635-59.

Aaberge, R., Colombino, U. and Strøm, S. (1999) Labour supply in Italy: an empirical analysis of joint household decisions with taxes and quantity constraints. *Journal of Applied Econometrics*, 14, 403-22.

Apps, P.F. (1994) Female labour supply, housework and family welfare. In R. Blundell, I. Preston and I. Walker (eds.) *The Measurement of Household Welfare* (pp.140-63). Cambridge: Cambridge University Press.

Apps, P.F. and Rees, R. (1996) Labour supply, household production and intra-family welfare distribution. *Journal of Public Economics*, 60, 199-219.

Apps, P.F. and Rees, R. (1997) Collective labor supply and household production. *Journal of Political Economy*, 105, 178-90.

Becker, G.S. (1965) A theory of the allocation of time. *Economic Journal*, 75, 493-517.

Bingley, P. and Walker, I. (1997) The labour supply, unemployment and participation of lone mothers in in-work transfer programmes. *Economic Journal*, 107, 1375-90.

Bingley, P. and Walker, I. (2001) Housing subsidies and work incentives in Great Britain. *Economic Journal*, 111, C86-C103.

Bingley, P., Lanot, G., Symons, E. and Walker, I. (1995) Child support reform and the labor supply of lone mothers in the United Kingdom. *Journal of Human Resources*, 30, 256-79.

Blundell, R., Chiappori, P.-A., Magnac, T. and Meghir, C. (1998) Collective labor supply: heterogeneity and nonparticipation. University College London.

Blundell, R., Duncan, A., McCrae, J. and Meghir, C. (2000) The labour market impact of the Working Families' Tax Credit. *Fiscal Studies*, 21, 75-104.

Blundell, R. and Hoynes, H. (2001), Has 'In-Work' Benefit Reform Helped the Labour Market. NBER Working Papers: 8546, National Bureau of Economic Research.

Bourguignon, F. and Chiappori, P.-A. (1994) The collective approach to household behaviour. In R. Blundell, I. Preston and I. Walker (eds.) *The Measurement of Household Welfare* (pp.70-85). Cambridge: Cambridge University Press.

Browning, M., Bourguignon, F., Chiappori, P.-A. and Lechene, V. (1994) Income and outcomes: a structural model of intrahousehold allocation. *Journal of Political Economy*, 102, 1067-96.

Burtless, G. and Hausman, J.A. (1978) The effect of taxation on labor supply: evaluating the Gary Negative Income Tax experiment. *Journal of Political Economy*, 86, 1103-30.

Callan, T. and Van Soest, A. (1996) Family labour supply and taxes in Ireland. Mimeo, Tilburg University.

Chiappori, P.-A. (1988) Rational household labour supply. *Econometrica*, 56(1), 63-89.

Creedy, J. and Duncan, A.S. (2002) Behavioural microsimulation with labour supply responses. *Journal of Economic Surveys*, 16, 1-39.

Creedy, J., Duncan, A.S., Harris, M. and Scutella, R. (2002) *Microsimulation Modelling of Taxation and The Labour Market: The Melbourne Institute Tax and Transfer Simulator*. Cheltenham: Edward Elgar.

Creedy, J., Kalb, G. and Kew, H. (2003) Flattening the effective marginal tax rate structure in Australia: policy simulations using the Melbourne Institute Tax and Transfer Simulator. *Australian Economic Review* 36(2), pp. 156-72.

Creedy, J., Kalb, G. and Scutella, R. (2004) Evaluating the Income Redistribution Effects of Tax Reforms in Discrete Hours Models. Forthcoming in *Research on Economic Inequality*, 12.

Duncan, A. and Harris, M.N. (2002) Simulating the effect of welfare reforms among sole parents in Australia. *Economic Record*, 78, 249-63.

Duncan, A. and Harris, M.N. (2002a) Intransigencies in the labour supply choice. Melbourne Institute Working Paper, no. 17/02.

Euwals, R. (2001) Female labour supply, flexibility of working hours, and job mobility. *Economic Journal*, 111, C120-34.

Euwals, R. and Van Soest, A. (1999) Desired and actual labour supply of unmarried men and women in the Netherlands. *Labour Economics*, 6, 95-118.

Fraker, T. and Moffitt, R. (1988) The effect of food stamps on labor supply: a bivariate selection model. *Journal of Public Economics*, 35, 25-56.

Gerfin, M. (1993) Simultaneous discrete choice model of labour supply and wages for married women in Switzerland. *Empirical Economics*, 18, 337-56.

Gerfin, M. and Leu, R.E. (2003) The impact of in-work benefits on poverty and household labour supply: a simulation study for Switzerland. Institute for the Study of Labor IZA Discussion Paper, no. 762.

Hagstrom, P.A. (1996) The food stamp participation and labor supply of married couples; an empirical analysis of joint decisions. *Journal of Human Resources*, 31, 383-403.

Hausman, J.A. (1979) The econometrics of labor supply on convex budget sets. *Economics Letters*, 3, 171-74.

Hausman, J.A. (1985) The econometrics of nonlinear budget sets. *Econometrica*, 53, 1255-82.

Hoynes, H.W. (1996) Welfare transfers in two-parent families: labor supply and welfare participation under AFDC-UP. *Econometrica*, 64, 295-332.

Kalb, G. (1999) Labour supply and welfare participation in Australian two-adult households: comparing 1986/1987 with 1994/1995. Working Paper No. BP-34, Centre of Policy Studies, Monash University, Australia.

Kalb, G. (2000) Accounting for involuntary unemployment and the cost of part-time work. Working Paper No. BP-35, Centre of Policy Studies, Monash University, Australia.

Keane, M. and Moffitt, R. (1998) A structural model of multiple welfare program participation and labor supply. *International Economic Review*, 39, 553-89.

Kooreman, P. and Kapteyn, A. (1987) A disaggregated analysis of the allocation of time within the household. *Journal of Political Economy*, 95, 223-41.

Kornstad, T. and Thoresen, T.O. (2002) A discrete choice model for labor supply and child care. Statistics Norway Research Department Discussion Paper.

Kornstad, T. and Thoresen, T.O. (2003) Means-testing the child benefit. *Review of Income and Wealth* (forthcoming)

McFadden, D. (1973) Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (ed.) *Frontiers of Econometrics* (pp.). New York: Academic Press.

McFadden, D. (1974) The measurement of urban travel demand. *Journal of Public Economics*, 3, 303-28.

MaCurdy, T., Green, D. and Paarsch, H. (1990) Assessing empirical approaches for analyzing taxes and labor supply. *Journal of Human Resources*, 25, 415-90.

Maddala, G.S. (1983) *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Moffitt, R. (1983) An economic model of welfare stigma. *American Economic Review*, 73, 1023-35.

Moffitt, R. (1986) The econometrics of piecewise-linear budget constraints: a survey and exposition of the maximum likelihood method. *Journal of Business and Economic Statistics*, 4, 317-28.

Scholz, J.K. (1996) In-work benefits in the United States: the Earned Income Tax Credit. *Economic Journal*, 106, 156-69.

Smith, P.A. (1997) The effect of the 1981 welfare reforms on AFDC participation and labor supply. Discussion Paper 1117-97, Institute for Research on Poverty, University of Wisconsin-Madison.

Tummers, M. and Woittiez, I. (1991) A simultaneous wage and labour supply model with hours restriction. *Journal of Human Resources*, 26, 393-423.

Van Soest, A. (1995) Structural models of family labor supply; a discrete choice approach. *Journal of Human Resources*, 30, 63-88.

Van Soest, A., Woittiez, I. and Kapteyn, A. (1990) Labor supply, income taxes, and hours restrictions in the Netherlands. *The Journal of Human Resources*, 25, 517-58.

Wales, T.J., and Woodland, A.D. (1977) Estimation of the allocation of time for work, leisure, and housework. *Econometrica*, 45, 115-32.

Woittiez, I. (1991) *Modelling and Empirical Evaluation of Labour Supply Behaviour: Studies in Contemporary Economics*. Berlin Heidelberg: Springer-Verlag:.

Zabalza, A., Pissarides, C. and Barton, M. (1980) Social security and the choice between full-time work, part-time work and retirement. *Journal of Public Economics*, 14, 245-76.

Appendix: An Iterative Solution Procedure

Suppose a function $F(x)$, with first derivative $f(x)$, needs to be maximised with regard to x. To find the maximum, the first order condition $f(x) = 0$ needs to be satisfied. Most iterative methods are based on some form of Newton's method. Consider finding the root of the equation $f(x) = 0$, where $f(x)$ takes the form shown in Figure 4. Take an arbitrary starting point, x_0 and draw the tangent, with slope $f'(x_0)$.
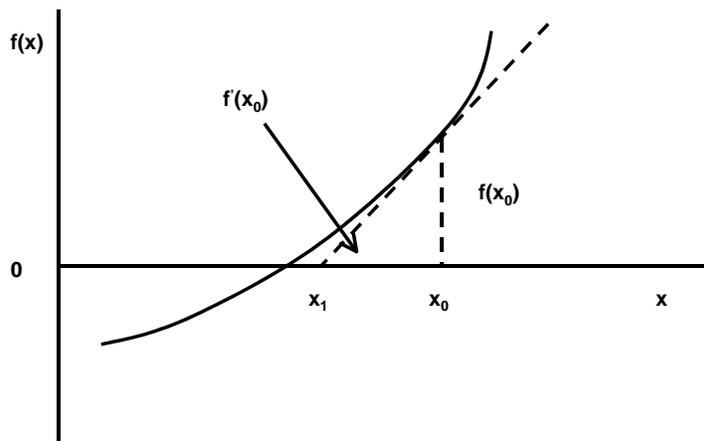

Figure 4: Newton's Method

By approximating the function by the tangent, the new value is given by the point of intersection of this tangent with the x axis, at x_1. It can be seen that selecting x_1 as the next starting point and drawing the tangent in this new point on f with slope $f'(x_1)$ leads quickly to the required root. From the triangle in Figure 4, it can be seen that:

$$f'(x_0) = \frac{f(x_0)}{x_0 - x_1} \tag{38}$$

Hence, starting from $I = 0$, the sequence of iterations follows:

$$x_{I+1} = x_I + \{-f'(x_I)\}^{-1} f(x_I) \tag{39}$$

60

until convergence is reached, when $x_{I+1} - x_I < \varepsilon$ with ε depending on the accuracy required. This clearly works best when the function is nice and smooth, and it is necessary to check (by picking different starting points) that there are not multiple roots, in which case convergence could be at a local rather than a global maximum. In addition, the second derivative $f'(x)$ needs to be negative in the maximum.