Creedy, John; Kalb, Guyonne

Working Paper

Discrete Hours Labour Supply Modelling: Specification, Estimation and Simulation

New Zealand Treasury Working Paper, No. 03/20

Provided in Cooperation with:
The Treasury, New Zealand Government

Suggested Citation: Creedy, John; Kalb, Guyonne (2003) : Discrete Hours Labour Supply Modelling: Specification, Estimation and Simulation, New Zealand Treasury Working Paper, No. 03/20, New Zealand Government, The Treasury, Wellington

This Version is available at:
https://hdl.handle.net/10419/205525

Discrete Hours Labour Supply Modelling: Specification, Estimation and Simulation

John Creedy and Guyonne Kalb

NEW ZEALAND TREASURY

WORKING PAPER 03/20

SEPTEMBER 2003

THE TREASURY
Kaitohutohu Kaupapa Rawa

NZ TREASURY WORKING PAPER 03/20

Discrete Hours Labour Supply Modelling: Specification, Estimation and Simulation

MONTH/YEAR

September 2003

AUTHORS

John Creedy
New Zealand Treasury
PO Box 3724
Wellington 6000
NEW ZEALAND

Email John.creedy@treasury.govt.nz]
Telephone 64-4-471-5009
Fax 64 4 473 1151

Guyonne Kalb
Melbourne Institute of Applied Economic and Social Research
University of Melbourne
Parkville, VIC 3010
AUSTRALIA

Email g.kalb@unimelb.edu.au
Telephone
Fax

ACKNOWLEDGEMENTS

We should like to thank Nathan McClellan and Jenny Williams for comments on an earlier draft of this paper

NZ TREASURY

New Zealand Treasury
PO Box 3724
Wellington 6000
NEW ZEALAND

Email information@treasury.govt.nz
Telephone 64-4-472 2733
Website www.treasury.govt.nz

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.

JEL CLASSIFICATION D10
C15
C25
C50

KEYWORDS Discrete hours labour supply, multinomial logit, maximum likelihood estimation, microsimulation

Table of Contents

List of Tables

List of Figures

Discrete Hours Labour Supply Modelling: Specification, Estimation and Simulation

1 Introduction

This paper provides an introduction to the basic analytics of discrete hours labour supply modelling.[1] Discrete hours models are popular in tax policy microsimulation, because it is relatively easy (compared to the continuous models) to incorporate taxation and social security details. To get to 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 parameters and individual information on incomes, wages and household composition are needed to calculate net incomes at all possible labour supply levels. Third, the model is estimated using information on individual 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 predict the effect of policy changes through simulation. 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. The discrete approach is perhaps more realistic, in that typically only a finite number of part-time or full-time working options are available.[2] It also substantially simplifies the nature of the budget set faced by each individual, who is assumed to face a fixed gross hourly wage. It is assumed that the same set of hours is available to 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.[3] 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.[4] With discrete hours models it is

[1] 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).

[2] 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.

[3] 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).

[4] This is further complicated in the case of couples and joint utility maximisation, where the budget constraint is three-dimensional.

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 (relating hours worked to net wage rates, non-wage incomes and a range of individual characteristics) is estimated, and then the 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, for example, 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. Indeed, empirical continuous hours models have found it extremely difficult to capture the complexities arising from supply behaviour under piece-wise linear constraints.

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 discrete-choice 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 crucially important probability distribution over the set of hours available for work. Section 3 provides a simple 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. Estimation of the parameters of specified preference functions, using the method of maximum likelihood, is considered in section 5. A numerical example of estimation is given in section 6. Alternative specifications of the model are discussed briefly in section 7. The use of discrete hours labour supply models in behavioural microsimulation is examined in section 8, where a numerical example of a tax reform is presented. Brief conclusions are in section 9.

2 The basic model

This section presents the basic model of utility maximisation and discusses the determination of the probability distribution of hours worked. Subsection 1 discusses the discrete choice framework, involving the introduction of a random term reflecting the difference between actual utility 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 and more formally in subsection 3. The measurement of labour supply elasticities in this framework is examined in subsection 4.

[5] A 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. A second generation of models examines 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.

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. The level of utility is determined by the amount of leisure and net income. Utility 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 (the total available time minus the time spent on leisure). 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 individual balances leisure and net income to obtain the highest utility possible, more leisure means less income and vice versa.

The utility associated with each hours level is denoted U_i^* and is a function of 'measured' utility $U(h_i|X)$ plus an 'error term', v_i, so that:[7]

$$U_i^* = U(h_i|X) + v_i$$

$$= U_i + v_i \qquad (1)$$

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 of course 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.[8] 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, $p(h = h_i) = p_i$, for $i = 1,...,n$, given assumptions about the distributions v_i.

2.2 Probability distributions

The framework, summarised by equation (1), is one in which there is 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 distributions of U_i^*, for $i = 1,2,3$, are shown in Figure 1, where in each case increasing utility involves moving upwards along each axis. The choice of any particular hours level is associated with

[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] 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.

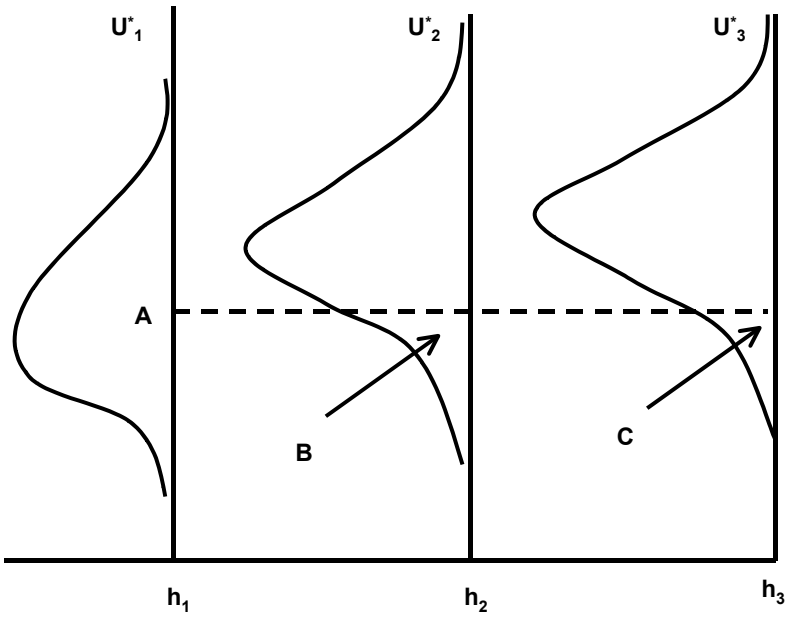
[8] In the next sections, emphasis is given to the case where the errors are independent and identically distributed.

'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.[9]

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^*.[10] 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.

Figure 1 – Three Hours Levels and Utility Distributions



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:

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

[9]The standard rule for independent probabilities is that *P*(*A* and *B*) = *P*(*A*) *P*(*B*).
[10]The appropriate combination of probabilities here follows the general rule that *P*(*A* or *B*) = *P*(*A*) + *P*(*B*).

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 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 context.

2.4 Labour supply elasticities

The structural basis of the discrete model means that there is no explicit labour supply function which depends on wage and other characteristics of the individual. This contrasts with the continuous hours approach where a supply function arises from utility maximisation subject to the budget constraint.[11] The estimated parameters are parameters 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 wage elasticity of labour supply can be applied in the discrete hours context.

An elasticity measure may be based on expected hours worked rather than a standard supply curve. 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.[12] 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 will 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

[11] 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.

[12] As shown in more detail in section 4.

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.

3 A numerical example of hours probabilities

This section shows how the probability distribution of an individual's hours of work is generated, 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):

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

Consider a situation in which there are just four hours levels of work 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, if the U_is were to represent utility precisely, h_3 would always be unambiguously chosen.

As above, 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 it is clear that h_4 is chosen in this case. It can be seen that, 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.

Table 2 – Conditional probabilities for hours level 3

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:

$$p_3 = \sum_{\substack{k=1 \\ k\neq 3}}^{4} P(U_3^* > U_j^*, \text{for all } j \neq 3 \,|\, v_3 = a_k)\; p(v_3 = a_k) \tag{6}$$

$$= \sum_{\substack{k=1 \\ k\neq 3}}^{4} \left(\prod_{j\neq 3} F(a_k + U_3 - U_j) \right) f(a_k) \tag{7}$$

$$= (0.032)(0.1) + (0.32)(0.3) + (0.80)(0.4) + (1.0)(0.2)$$

$$= 0.619 \tag{8}$$

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. The multinomial logit 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 state 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:

$$p_i = \int_{-\infty}^{+\infty} \left\{ \prod_{j \neq i} F\left(v_i + U_i - U_j\right) \right\} f(v_i)\, dv_i \tag{9}$$

Essentially, the expression in (9) 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, $F(.)$, it cannot be expected that an arbitrary choice of $f(v)$ will 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:

$$f(v) = e^{-v} e^{-e^{-v}}$$

$$= \exp\left(-v - e^{-v}\right) \tag{10}$$

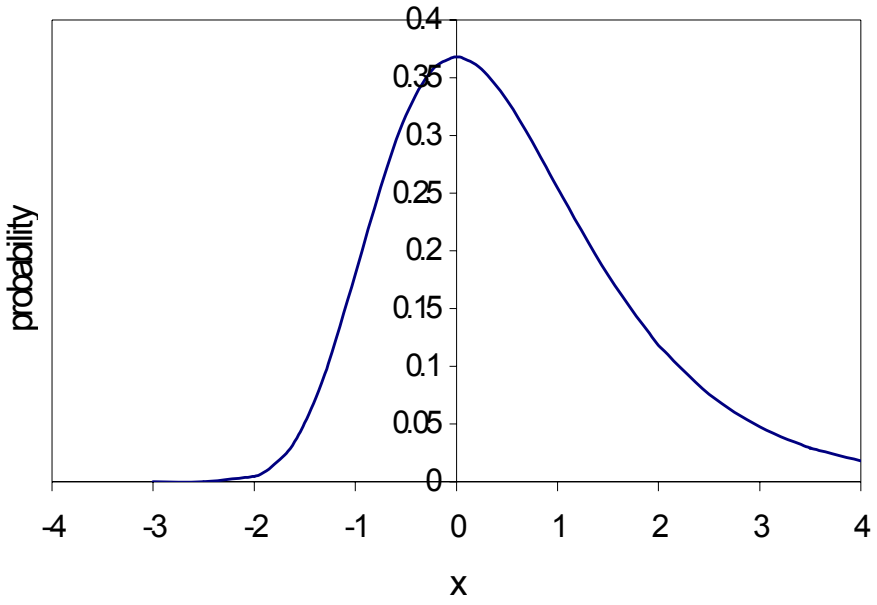
for which the distribution function is:

$$F(v) = e^{-e^{-v}} \tag{11}$$

The choice of this 'thin-tailed' distribution has the obvious advantage that no further parameters need to be estimated.[13] This is known as an Extreme (Maximum) Value Type I distribution, which is often abbreviated to 'extreme value' distribution.[14] It is highly tractable in the present context. These qualities have generally been (implicitly) 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 $-\ln(\ln 2)$. 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.

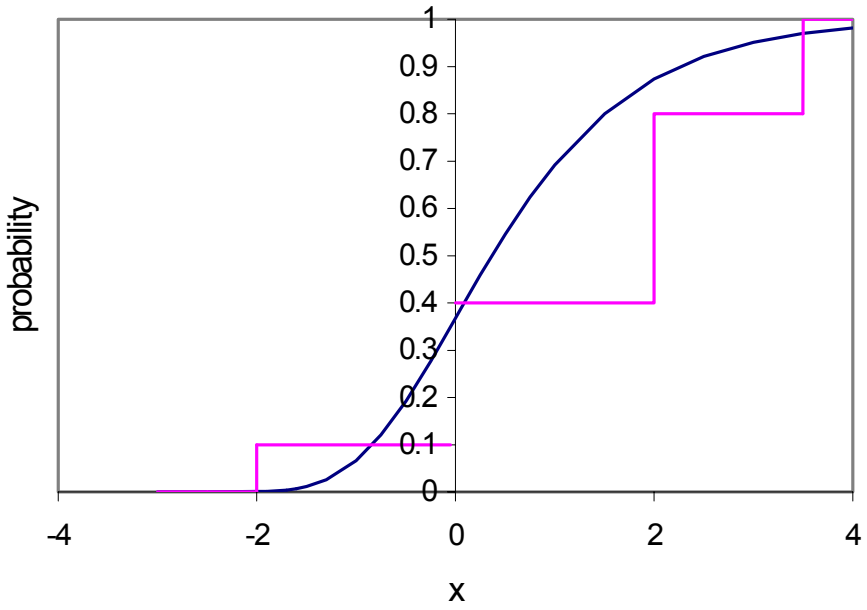
Figure 2 – Extreme value probability density function



[13] 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).

[14] This is because it has been found useful in many applications involving extreme values. If a process generating values of 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.

Figure 3 – Extreme value cumulative distribution function and the discrete distribution from Table 1



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

$$\exp\left(-v_i - e^{-v_i}\right)\prod_{j\neq i}\exp\left(-e^{-v_i}e^{-U_i+U_j}\right) \tag{12}$$

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 (12) becomes:

$$\exp\left[-v_i - e^{-v_i}\left(1 + e^{-U_i}\sum_{j\neq i}e^{U_j}\right)\right] \tag{13}$$

Furthermore, $1 + e^{-U_i}\sum_{j\neq i}e^{U_j} = e^{U_i}e^{-U_i} + e^{-U_i}\sum_{j\neq i}e^{U_j} = e^{-U_i}\sum_{j=1}^{n}e^{U_j} = e^{\lambda_i}$, say. Hence (13) can be rewritten more succinctly as $\exp\left[-v_i - e^{-\left(v_i-\lambda_i\right)}\right]$. Thus:

$$p_i = \int_{-\infty}^{+\infty}\exp\left[-v_i - e^{-\left(v_i-\lambda_i\right)}\right]dv_i \tag{14}$$

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

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

$$= e^{-\lambda_i} \int_{-\infty}^{+\infty} \exp\left[-v_i' - e^{-v_i'}\right] dv_i'$$

$$= e^{-\lambda_i} \int_{-\infty}^{+\infty} f\left(v_i'\right) dv_i'$$

$$= e^{-\lambda_i}$$

$$= \frac{e^{U_i}}{\sum_{j=1}^{n} e^{U_j}} \tag{15}$$

In this special case, the probability distribution of hours of work for an individual depends in a very simple way on the measured utility levels associated with each hours level.[15] The discrete choice model flowing from the assumption of an extreme value distribution is called a multinomial logit model.[16]

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 (15) gives the probabilities 0.005, 0.056, 0.686, and 0.253.

5 Parameter estimation

The previous sections have examined the discrete choice model underlying an individual's labour supply behaviour. The basic assumptions are that individuals maximize their utility and that utility 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 method of maximum likelihood. 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 1 and its maximisation is considered in subsection 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 k th person; denote this by h_{i_k}, so that i_k indicates the chosen value of i (the hours index) for person k. Consistent with this notation, the probability of selecting this hours level is p_{i_k}

[15] 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.

[16] For an extensive comparison of alternative discrete choice models, see Maddala (1983).

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:

$$p_{i_k} = P\left(U^*_{i_k,k} \geq U^*_{j,k} \text{ for all } j\right) \tag{16}$$

Thus when all v_i are assumed to follow the extreme value distribution discussed in Section 1, the probability associated with the optimal hours chosen by person k is expressed as:

$$p_{i_k} = \frac{e^{U_{i_k,k}}}{\sum_{j=1}^{n} e^{U_{j,k}}} \tag{17}$$

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:

$$P(h_{i_1},...,h_{i_M}) = p_{i_1} p_{i_2} ... p_{i_M} = \prod_{k=1}^{M} \frac{e^{U_{i_k,k}}}{\sum_{j=1}^{n} e^{U_{j,k}}} \tag{18}$$

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.

The situation facing researchers is that they 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.[17]

The probability in (18) 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 (18) 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 necessary 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.

[17]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 using estimated wage functions. An alternative is to estimate a joint wage and labour supply model (see for example, Gerfin, 1993).

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 (18) can be rewritten as:

$$L(\beta_1,...,\beta_S) = \prod_{k=1}^{M} \frac{e^{U_{i_k,k}}}{\sum_{j=1}^{n} e^{U_{j,k}}} \tag{19}$$

where $L(\beta_1,...,\beta_S)$, a function of the unknown parameters (for a given sample of observed hours worked), is referred to as the Likelihood Function. Here the (fixed) parameters are effectively treated as if they were variables. The estimates, $\hat{\beta}_1,...,\hat{\beta}_S$ 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:

$$\log L = \sum_{k=1}^{M} \left[U_{i_k,k} - \log\left(\sum_{j=1}^{n} e^{U_{j,k}} \right) \right] \tag{20}$$

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:[18]

$$\frac{\partial \log L}{\partial \beta_\ell} = 0 \text{ for all } \ell = 1,...,S \tag{21}$$

Differentiation of (20) gives:[19]

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

$$= \sum_{k=1}^{M} \left[\frac{\partial U_{i_k,k}}{\partial \beta_\ell} - \sum_{j=1}^{n} \left(\frac{e^{U_{j,k}}}{\sum_{j=1}^{n} e^{U_{j,k}}} \right) \frac{\partial U_{j,k}}{\partial \beta_\ell} \right] \tag{22}$$

In considering the terms in (22) 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.

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

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

It is of interest to rewrite the first-order conditions, using (21) and (22) as giving, for all ℓ:

$$\sum_{k=1}^{M}\left[\frac{\partial U_{i_k,k}}{\partial \beta_\ell}-\sum_{j=1}^{n} p_{j,k}\frac{\partial U_{j,k}}{\partial \beta_\ell}\right]=0 \tag{23}$$

This has the simple 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 any practical help in trying to solve the highly nonlinear set of equations.

The solution (the set of maximum likelihood estimates for all β_ℓ s) can be obtained using numerical methods involving a sequence of iterations which lead efficiently from an arbitrary starting point to the solution. A discussion of Newton's method, which is often used to maximise functions, can be found in the appendix.

The iterative method involves repeatedly solving the following matrix equation, where $\beta^{[I]}$ denotes the vector of parameters in the I th iteration:

$$\beta^{[I+1]}=\beta^{[I]}+\left[-\frac{\partial^2 \log L}{\partial \beta_\ell \partial \beta_s}\right]^{-1}_{\beta^{[I]}}\left[\frac{\partial \log L}{\partial \beta_\ell}\right]_{\beta^{[I]}} \tag{24}$$

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. The application of Newton's method in the present context therefore requires the second derivatives of the likelihood function. Differentiating (22) again with respect to parameter β_s, gives:

$$\frac{\partial^2 \log L}{\partial \beta_\ell \partial \beta_s}=\sum_{k=1}^{M}\left[\frac{\partial^2 U_{i_k,k}}{\partial \beta_\ell \partial \beta_s}-\sum_{j=1}^{n}\left\{p_{j,k}\frac{\partial^2 U_{j,k}}{\partial \beta_\ell \partial \beta_s}+\frac{\partial U_{j,k}}{\partial \beta_\ell}\frac{\partial p_{j,k}}{\partial \beta_s}\right\}\right] \tag{25}$$

where

$$\frac{\partial p_{j,k}}{\partial \beta_s}=p_{j,k}\left[\frac{\partial U_{j,k}}{\partial \beta_s}-\sum_{t=1}^{n} p_{t,k}\frac{\partial U_{t,k}}{\partial \beta_s}\right] \tag{26}$$

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 the example has few individuals and a simple utility specification, the general approach is no different in a more realistic example. 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. Hence, all individuals have the same utility function with the same parameters, and this takes the form:

$$U^* = \beta_h h + \beta_y y + v \tag{27}$$

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 y at the hours points h and so optimal hours are located at different points and, secondly, the error term v introduces random differences in utility caused by unobserved factors. In this simple linear form, 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, so 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 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, given 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.[20]

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

Assuming for simplicity that there are no income taxes or benefit payments, the net income is simply equal to gross income. Substitution of these values, along with (27), into the first-order conditions in (22) give:[21]

$$\left[\frac{\partial \ln L}{\partial \beta_h}\right]_{\beta} = -\frac{20\exp(\beta_h 20 + \beta_y 80) + 40\exp(\beta_h 40 + \beta_y 160)}{1 + \exp(\beta_h 20 + \beta_y 80) + \exp(\beta_h 40 + \beta_y 160)}$$

$$+20 - \frac{20\exp(\beta_h 20 + \beta_y 160) + 40\exp(\beta_h 40 + \beta_y 320)}{1 + \exp(\beta_h 20 + \beta_y 160) + \exp(\beta_h 40 + \beta_y 320)}$$

$$+40 - \frac{20\exp(\beta_h 20 + \beta_y 200) + 40\exp(\beta_h 40 + \beta_y 400)}{1 + \exp(\beta_h 20 + \beta_y 200) + \exp(\beta_h 40 + \beta_y 400)}$$

$$= 0 \tag{28}$$

[20] If allowance were made for other characteristics and given the error term, this would not necessarily always be the case; some low-wage individuals work long hours, and *vice versa*.

[21] This specification for ν 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.

$$\left[\frac{\partial \ln L}{\partial \beta_y}\right]_{\beta} = -\frac{80\exp(\beta_h 20+\beta_y 80)+160\exp(\beta_h 40+\beta_y 160)}{1+\exp(\beta_h 20+\beta_y 80)+\exp(\beta_h 40+\beta_y 160)}$$

$$+160-\frac{160\exp(\beta_h 20+\beta_y 160)+320\exp(\beta_h 40+\beta_y 320)}{1+\exp(\beta_h 20+\beta_y 160)+\exp(\beta_h 40+\beta_y 320)}$$

$$+400-\frac{200\exp(\beta_h 20+\beta_y 200)+400\exp(\beta_h 40+\beta_y 400)}{1+\exp(\beta_h 20+\beta_y 200)+\exp(\beta_h 40+\beta_y 400)}$$

$$=0 \qquad (29)$$

The result is two nonlinear equations in the two unknowns β_y and β_h. Using an iterative solution procedure, as described in section 2, the maximum likelihood estimates were found to be $\beta_h = -15.41$ and $\beta_y = 1.93$.[22]

Consider the wage elasticity of labour supply for person 2, defined as in subsection 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 Table 3 are used, with the parameter estimates, to obtain the utilities corresponding to each hours point, by appropriate substitution in $U_{i2} = \beta_h h_i + \beta_y y_{i2}$. 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 a very high elasticity of 43.

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. First, the form of utility functions is examined in subsection 1. Allowance for participation in welfare programmes, often described in terms of the 'take-up' of benefits, is examined in subsection 2. Alternative ways in which allowance may be made for individuals' personal characteristics are discussed in subsection 3. The effect of characteristics of a particular discrete hours point is described in subsection 4. Finally, alternatives to the use of the extreme value distribution are briefly discussed in subsection 5.

[22]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.

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 extension in applied work is the quadratic utility function:[23]

$$U = \beta_y y + \beta_h h + \alpha_y y^2 + \alpha_h h^2 + \alpha_{yh} yh \tag{30}$$

where the marginal utility of income is:

$$\frac{\partial U}{\partial y} = \beta_y + 2\alpha_y y + \alpha_{yh} h \tag{31}$$

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:[24]

$$U = \beta_y \ln y + \beta_l \ln(T-h) + \alpha_y (\ln y)^2 + \alpha_l (\ln(T-h))^2 + \alpha_{yl} \ln y \ln(T-h) \tag{32}$$

where the marginal utility of income is:

$$\frac{\partial U}{\partial y} = \frac{\beta_y + 2\alpha_y \ln y + \alpha_{yl} \ln(T-h)}{y} \tag{33}$$

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 (if α_{yh} is negative or α_{yl} is positive) or substitutability (if α_{yh} is positive or α_{yl} is negative) of income and leisure. For example, the value of income may increase if more 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 is cumbersome in maximum likelihood estimation.[25] In discrete hours labour supply modelling, it is sufficient to check for quasi-concavity after estimation, which is a straightforward check of two necessary conditions.[26]

The quadratic and translog utility function 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

[23] Examples of the use of this can be found in, for example, Keane and Moffitt (1998).
[24] This has been used by, for example, Van Soest (1995).
[25] See MaCurdy, Green and Paarsch (1990).
[26] The two conditions are discussed by Van Soest (1995). The first is the basic requirement that utility increases with income. The second condition is more complicated but straightforward to check.

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 break down consumption into shared and private goods or to construct a sharing rule for income.[27] 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:

$$U = \beta_y y + \beta_m h_m + \beta_w h_w + \alpha_y y^2 + \alpha_m h_m^2 + \alpha_w h_w^2 + \alpha_{ym} y h_m + \alpha_{yw} y h_w + \alpha_{wm} h_w h_m \tag{34}$$

where the index m denotes hours and parameters of the male and the index w denotes hours and parameters of the female, and y represents joint income. The parameter α_{wm} 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.[28] 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. In these models, it is expected that disutility is attached to participation in welfare. This disutility could be caused for example by the costs of applying for welfare. 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 of being on welfare, 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 and popular way of adding welfare participation to the utility function is through the addition of a dummy variable for participation.[29] For example, $d_w = 1$ 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 would therefore be:

$$U = \beta_y y + \beta_h h + \alpha_y y^2 + \alpha_h h^2 + \alpha_{yh} y h + \phi d_w \tag{35}$$

The participation parameter can be made dependent on individual characteristics in the same way as for the preference 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

[27] 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).
[28] Moffitt (1983) introduced this idea.
[29] Examples of this approach can be found in Fraker and Moffitt (1988), Hoynes (1996), Hagstrom (1996), Smith (1997), Keane and Moffitt (1998), Kalb (1999, 2000).

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.[30]

7.3 Personal characteristics

Consider again the simple linear utility function:

$$U^* = \beta_h h + \beta_y y + v \tag{36}$$

It is straightforward to extend this to make the preference parameters dependent on personal and household characteristics. Characteristics such as education, number and age of children or an individual's own age are likely to influence the preference for work and income. Including these characteristics in the preference for work parameter, the utility function could be presented as follows:

$$U^* = \{\beta_{h1} + \beta_{h2} age + \beta_{h3} d\} h + \beta_y y + v \tag{37}$$

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 easily be examined. Expected labour supply can be calculated for two individuals who are exactly the same except for the characteristic of interest. There is thus scope for a wide range of elasticities.

The approach reflected in equation (37) 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:

$$\beta_h = \beta_{h1} + \beta_{h2} age + \beta_{h3} d + \varepsilon \tag{38}$$

This introduces an additional error term, ε, which is typically assumed to be normally distributed. This addition complicates the method of estimation somewhat, such that the method of simulated maximum likelihood is required. However, estimation of such models, including correlated error terms in the different preference terms, remains fairly straightforward using this method.[31]

[30] See Bingley and Walker (1997, 2001), who estimate 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.
[31] See for example Van Soest (1995).

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 only three discrete hours points. This contrasts with around ten hours points for each individual, when using the extreme value distribution.[32] 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 supply and welfare participation.

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.[33] 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.[34] 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.[35] Third, a parameter measuring the fixed cost of working can be subtracted from net income in a quadratic utility function.[36] 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.[37] A final example is the approach where an adaptation of the multinomial logit model allows for captivity at particular discrete hours points.[38] 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 characteristics; this may, for example, be expected at the standard full-time 40-hours point.

[32] This remains possible even when labour supply is estimated jointly for couples.

[33] Euwals (2001) shows 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.

[34] See for example Van Soest (1995), Callan and Van Soest (1996) or Kalb (2000).

[35] See for example, Woittiez (1991) or Euwals and Van Soest (1999). The first uses the hours restrictions as a way of specifying a discrete model, that is, the discrete points have positive probability of being in the choice set of the individual. The latter takes desired labour supply as given and examines the probability of obtaining job offers at the different hours points separately.

[36] See for example, Duncan and Harris (2002).

[37] See Aaberge, Dagsvik and Strøm (1995), Aaberge, Colombino and Strøm (1999) and Kornstad and Thoresen (2002, 2003).

[38] See Duncan and Harris (2002a).

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.

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.[39] 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.

[39] 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 $U(20,y_{20}) = U(40,y_{40})$.

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. That is the distribution of the differences in utility between nonparticipation 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 would be 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.[40] Although the intuition behind this model is simple, estimation of the model is difficult, particularly for large M.

8 Tax reforms and simulations

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. 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. 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.[41] 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.

First, subsection 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

[40] 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. Input 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.

[41] These contrast with, for example, changes in 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 transfer system, but are difficult to accommodate in microsimulation.

each individual, conditional on them being at their observed pre-reform hours, is also described. Secondly, subsection 2 provides a small numerical example of a tax reform, using the three hypothetical individuals used in the illustration of maximum likelihood estimation.

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.[42] 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 individual, 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 is drawn.

These sets of error terms that resulted in the observed labour supply are then used to compute a distribution of labour supply after a specified reform.[43] 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 $U = -15.41h + 1.93y$. 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.[44] Table 4 presents the income and utility at the discrete hours points for all three individuals before and after the reform.

[42]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).

[43]The more error terms that are drawn, the more accurate is the computed distribution, especially for those points with low probability.

[44]This is sometimes described as a basic income - flat tax structure, or a social dividend scheme, or a negative income tax.

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

From the table it is clear that the introduction of the tax system has made work much less attractive. Adding 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. 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 this draw, after the reform has been introduced, results in utility levels post-reform, indicated by U^{*1}. From the utility levels in Table 4, it is clear that 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 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.

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	-1.070	1.361	0.178	2.997	3.491	0.217	1.176	1.026	2.426
U^{*0}	-1.070	-152.439	-307.422	2.997	4.091	1.417	1.176	78.826	158.026
U^{*1}	27.880	-154.369	-340.232	31.947	-28.719	-93.153	30.126	30.576	32.576
v_2	-0.805	0.437	-0.012	0.777	0.241	-1.416	0.907	-0.781	5.678
U^{*0}	-0.805	-153.363	-307.612	0.777	0.841	-0.216	0.907	77.019	161.278
U^{*1}	28.145	-155.293	-340.422	29.727	-31.969	-94.786	29.857	28.769	35.828
v_3	0.233	-0.742	0.037	-0.168	1.285	0.080	0.801	1.232	0.992
U^{*0}	0.233	-154.542	-307.563	-0.168	1.885	1.280	0.801	79.032	156.592
U^{*1}	29.183	-156.472	-340.373	28.782	-30.925	-93.290	29.751	30.782	31.142
v_4	2.554	2.402	-0.022	-0.638	1.635	0.522	2.069	1.249	0.456
U^{*0}	2.554	-151.398	-307.622	-0.638	2.235	1.722	2.069	79.049	156.056
U^{*1}	31.504	-153.328	-340.432	28.312	-30.575	-92.848	31.019	30.799	30.606
v_5	-0.019	0.656	1.257	0.712	3.741	2.412	-0.715	-0.400	-0.329
U^{*0}	-0.019	-153.144	-306.343	0.712	4.341	3.612	-0.715	77.400	155.271
U^{*1}	28.931	-155.074	-339.153	29.662	-28.469	-90.958	28.235	29.150	29.821
v_6	0.062	1.628	-1.269	0.113	2.428	-0.412	-1.243	-0.673	-0.535
U^{*0}	0.062	-152.172	-308.869	0.113	3.028	0.788	-1.243	77.127	155.065
U^{*1}	29.012	-154.102	-341.679	29.063	-29.782	-93.782	27.707	28.877	29.615
v_7	-0.626	1.079	-0.550	1.196	0.844	-1.501	1.771	1.518	2.311
U^{*0}	-0.626	-152.721	-308.150	1.196	1.444	-0.301	1.771	79.318	157.911
U^{*1}	28.324	-154.651	-340.960	30.146	-31.366	-94.871	30.721	31.068	32.461
v_8	0.136	1.233	0.174	-0.507	1.855	1.036	-1.346	-0.555	1.123
U^{*0}	0.136	-152.567	-307.426	-0.507	2.455	2.236	-1.346	77.244	156.723
U^{*1}	29.086	-154.497	-340.236	28.443	-30.355	-92.334	27.604	28.995	31.273
v_9	2.745	-0.530	0.363	0.163	1.044	-0.216	0.633	0.433	-0.695
U^{*0}	2.745	-154.330	-307.237	0.163	1.644	0.984	0.633	78.233	154.905
U^{*1}	31.695	-156.260	-340.047	29.113	-31.166	-93.586	29.583	29.983	29.455
v_{10}	1.730	-0.330	-1.190	-1.240	1.479	-0.861	-0.497	0.187	0.229
U^{*0}	1.730	-154.130	-308.790	-1.240	2.079	0.339	-0.497	77.987	155.829
U^{*1}	30.680	-156.060	-341.600	27.710	-30.731	-94.231	28.453	29.737	30.379

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

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 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 extremely 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 probability of 10 per cent of moving out of the labour force and the probability of reducing labour supply 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

The predicted probability of person 3 being in zero hours, 20 hours and 40 hours is 15.4 per cent, 28.1 per cent and 56.5 per cent respectively.[45] 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 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 the approximation becomes more accurate.[46]

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. 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, because for the other two persons nearly all possible draws of the error term result in the correct 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.

[45] These probabilities are calculated by computing $P(h_1) = \frac{\exp(U_{1,1})}{\exp(U_{1,1})+\exp(U_{2,1})+\exp(U_{3,1})} = \frac{\exp(28.9)}{\exp(28.9)+\exp(29.5)+\exp(30.2)} = 0.154$, and similar expressions for the other hours points.

[46] For example for 20 draws, the percentages at 0, 20 and 40 hours are 20, 35 and 45 per cent respectively.

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

9 Conclusions

This paper has provided an introduction to the basic analytics of discrete hours labour supply modelling. Special attention was given to model specification, estimation and microsimulation. The paper has given several numerical examples to illustrate the more technical exposition of the methodologies used in this research field. It is suggested that the approach offers much potential for further interesting and valuable applications and extensions.

Several developments 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.

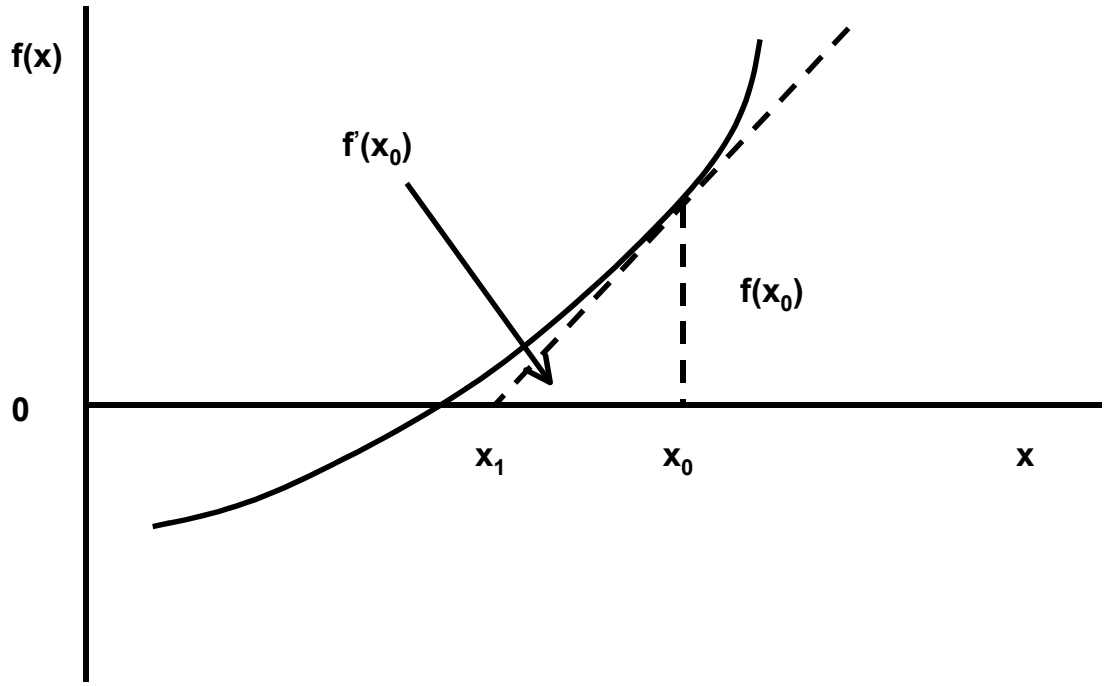
One area related to the discussion in this paper, that has received little attention in the literature so far, is concerned with the evaluation of simulation outcomes. When using discrete choice labour supply models in simulation, the outcomes of analyses are probabilistic in nature. Measures of welfare, inequality or poverty which can deal with these probabilistic outcomes need further development.[47]

[47] Creedy, Kalb and Scutella (2003) propose an approach for calculating inequality and poverty measures in a discrete choice microsimulation setting.

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:

$$f'(x_0)=\frac{f(x_0)}{x_0-x_1} \tag{39}$$

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

$$x_{I+1}=x_I+\left\{-f'(x_I)\right\}^{-1}f(x_I) \tag{40}$$

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.

In the present context, Newton's method is easily adapted to deal with a vector of parameters. An iterative method involves repeatedly solving the following matrix equation, where $x^{[I]}$ now denotes the vector of parameters in the I th iteration:

$$x^{[I+1]} = x^{[I]} + \left[-\frac{\partial^2 F}{\partial x_\ell \partial x_s} \right]^{-1}_{x^{[I]}} \left[\frac{\partial F}{\partial x_\ell} \right]_{x^{[I]}} \tag{41}$$

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.

References

[1] Aaberge, R., J.K. Dagsvik and S. Strøm (1995) Labor supply responses and welfare effects of tax reforms. *Scandinavian Journal of Economics*, 97, 635-659.

[2] Aaberge, R., U. Colombino and S. Strøm (1999) Labour supply in Italy: an empirical analysis of joint household decisions with taxes and quantity constraints. *Journal of Applied Econometrics*, 14, 403-422.

[3] Apps, P.F. (1994) Female labour supply, housework and family welfare. In *The Measurement of Household Welfare* (ed. by R. Blundell, I. Preston and I. Walker). Cambridge: Cambridge University Press.

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

[5] Apps, P.F. and R. Rees (1997) Collective labor supply and household production. *Journal of Political Economy*, 105, 178-190.

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

[7] Bingley, P. and I. Walker (1997) The labour supply, unemployment and participation of lone mothers in in-work transfer programmes. *Economic Journal*, 107, 1375-1390.

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

[9] Bingley, P., G. Lanot, E. Symons and I. Walker (1995) Child support reform and the labor supply of lone mothers in the United Kingdom. *Journal of Human Resources*, 30, 256-279.

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

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

[12] Bourguignon, F. and P.-A. Chiappori (1994) The collective approach to household behaviour. In *The Measurement of Household Welfare* (ed. by R. Blundell, I. Preston and I. Walker). Cambridge: Cambridge University Press.

[13] Browning, M., F. Bourguignon, P.-A. Chiappori and V. Lechene (1994) Income and outcomes: a structural model of intrahousehold allocation. *Journal of Political Economy*, 102, 1067-1096.

[14] Burtless, G. and J.A. Hausman (1978) The effect of taxation on labor supply: evaluating the Gary Negative Income Tax experiment. *Journal of Political Economy*, 86, 1103-1130.

[15] Callan, T. and A. van Soest (1996) Family labour supply and taxes in Ireland, mimeo, Tilburg University.

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

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

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

[19] Creedy, J., Kalb, G. and H. Kew (2003) Flattening the effective marginal tax rate structure in Australia: policy simulations using the Melbourne Institute Tax and Transfer Simulator. *Australian Economic Review* (forthcoming).

[20] Creedy, J., Kalb, G. and R. Scutella (2003) Evaluating the Income Redistribution Effects of Tax Reforms in Discrete Hours Models. mimeo, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

[21] Duncan, A. and M.N. Harris (2002) Simulating the effect of welfare reforms among sole parents in Australia. *Economic Record*, 78, 249-263.

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

[23] Euwals, R. (2001) Female labour supply, flexibility of working hours, and job mobility. *Economic Journal*, 111, C120-C134.

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

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

[26] Gerfin, M. (1993) Simultaneous discrete choice model of labour supply and wages for married women in Switzerland. *Empirical Economics*, 18, 337-356.

[27] Gerfin, M. and R.E. Leu (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.

[28] 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.

[29] Hausman, J.A. (1979) The econometrics of labor supply on convex budget sets. *Economics Letters*, 3, 171-174.

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

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

[32] 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.

[33] 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.

[34] Keane, M. and R. Moffitt (1998) A structural model of multiple welfare program participation and labor supply. *International Economic Review*, 39, 553-589.

[35] Kooreman, P. and A. Kapteyn (1987) A disaggregated analysis of the allocation of time within the household. *Journal of Political Economy*, 95, 223-241.

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

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

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

[39] McFadden, D. (1974) The measurement of urban travel demand. *Journal of Public Economics*, 3, 303-328.

[40] MaCurdy, T., D. Green, and H. Paarsch (1990) Assessing empirical approaches for analyzing taxes and labor supply. *Journal of Human Resources*, 25, 415-490.

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

[42] Moffitt, R. (1983) An economic model of welfare stigma. *American Economic Review*, 73, 1023-1035.

[43] 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, 317-328.

[44] Scholz, J.K. (1996) In-work benefits in the United States: the Earned Income Tax Credit. *Economic Journal*, 106, 156-169.

[45] 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.

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

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

[48] Van Soest, A., I. Woittiez and A. Kapteyn (1990) Labor supply, income taxes, and hours restrictions in the Netherlands. *The Journal of Human Resources*, 25, 517-558.

[49] Wales, T.J., and A.D. Woodland (1977) Estimation of the allocation of time for work, leisure, and housework. *Econometrica*, 45, 115-132.

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

[51] Zabalza, A., C. Pissarides and M. Barton (1980) Social security and the choice between full-time work, part-time work and retirement. *Journal of Public Economics*, 14, 245-276.