SNORRe

Statistics Norway's Open Research Repository

Title:	Evaluating alternative representations of the choice sets in models of labour supply
Author:	Aaberge, Rolf Colombino, Ugo Wennemo, Tom
Version:	**Authors Submitted Version / Pre-print** **This is the pre-peer reviewed version of the following article: Journal of Economic Surveys, vol. 23, 3, 586-612, which has been published in final form at: doi: 10.1111/j.1467-6419.2008.00573.x**
Note:	
Publisher:	Wiley-Blackwell
DOI:	http://dx.doi.org/ **10.1111/j.1467-6419.2008.00573.x**

Author's website: http://www.ssb.no/english/research/people/roa/index.html

Evaluating Alternative Representations of the Choice Sets in Models of Labour Supply

R. Aaberge, Statistics Norway, Oslo, Norway
U. Colombino, Department of Economics, Turin, Italy
T. Wennemo, Statistics Norway, Oslo, Norway

Corresponding author: Ugo Colombino, Department of Economics, Via Po 53, 124 Torino (IT); Phone: +390116703860; Fax: +390116703895; e-mail: ugo.colombino@unito.it.

Evaluating Alternative Representations of the Choice Sets in Models of Labour Supply

Abstract

During the last two decades, the discrete-choice modelling of labour supply decisions has become increasingly popular, starting with Aaberge et al. (1995) and van Soest (1995). Within the literature adopting this approach there are however two potentially important issues that so far have not been given the attention they might deserve. A first issue concerns the procedure by which the discrete alternatives are selected to enter the choice set. For example van Soest (1995) chooses (not probabilistically) a set of fixed points identical for every individual. This is by far the most widely adopted method. By contrast, Aaberge et al. (1995) adopt a sampling procedure suggested by McFadden (1978) and also assume that the choice set may differ across the households. A second issue concerns the availability of the alternatives. Most authors assume all the values of hours-of-work within some range are equally available. At the other extreme, some authors assume only two or three alternatives (e.g. non-participation, part-time and full-time) are available for everyone. By contrast, Aaberge et al. (1995) account for the fact that not all the hour opportunities are equally available to everyone specifying a probability density function of opportunities for each individual. The discrete choice set used in the estimation is built by sampling from that individual-specific density function. In this paper we explore by simulation the implications of: (i) the procedure used to build the choice set (fixed alternatives *versus* sampled alternatives); (ii) accounting or not accounting for a different availability of alternatives. The results of the evaluation performed in this paper show that the way the choice set is represented has little impact on the fitting of observed values, but a more significant and important impact on the out-of-sample prediction performance. Thus, the treatment of the choice sets might have a crucial effect on the result of policy evaluations.

Key words: Discrete choice models, Random utility models, Choice set specification, Labour supply, Prediction performance

JEL: C51, C52, H31

1. Introduction

The idea of modelling labour supply decisions as discrete choices has become more and more popular during the last two decades. In this paper we examine, through a simulation exercise, an issue that has received much less attention than it might deserve: the implications of alternative methods of representing the choice set within the discrete choice approach.

The discrete choice approach has gained a prominent position as the outcome of a process aimed at solving or circumventing some theoretical and computational problems to be faced with in micro-econometric research when analyzing choices subject to complicated constraints. The beginning of this process might be traced back to the late 60s and early 70s, when a strong interest emerged in designing and evaluating various welfare and "anti-poverty" programs. These policies introduce complications (non-linearities, non-convexities) into the budget sets faced by the target population, which are hard to deal with within the standard framework based on demand (or supply) functions. Perhaps Heckman (1974) represents the first contribution that fully clarifies the issue. The policy problem addressed is the evaluation of a child-related welfare policy that introduces significant complications in the budget set. Heckman observed that in order to make such evaluation one has to estimate the preferences as separated from the constraints: " *The essence of the problem involves utility comparisons between two or more discrete alternatives. Such comparisons inherently require information about consumer preferences in a way not easily obtained from ordinary labor-supply functions*" (Heckman 1974, page S136). Moreover "...*the ability to make* ...(the separation between preferences and constraints) ... *is less important if we are willing to make the conventional assumption that wage rates are independent of hours of work ... but becomes quite important when we acknowledge the existence of progressive taxation, welfare regulations, and time and money costs of work*" (Heckman 1974, page S142). In that paper Heckman proposed a particular method of identifying indifference curves as envelopes of tangents. In the same period, J. Hausman and various co-authors addressed essentially the same problem and proposed a method specifically appropriate for piece-wise linear budget constraints (e.g. Hausman, 1979). These contributions work through the implications of the Kuhn-Tucker conditions associated to the maximization of utility subject to inequality constraints. The solution can be located in different ranges of values along the budget constraint. Corresponding to each possible range of values there is a condition involving the preference parameters. Choosing a convenient stochastic specification, we can express the probability that those various conditions alternatively hold, write down the sample likelihood and estimate the preference parameters. Useful presentations of this class of methods have been provided by Moffit (1986), Blomquist (1988) and Blundell and MaCurdy (2000).

The method proposed by Heckman as well as the method proposed by Hausman and co-authors are in principle fairly general but might in practice turn out to be not so easily applicable to problems that are more complicated than those for which they were originally exemplified. More specifically, as far as the Hausman and co-authors's approach is concerned, the experience suggests that the method presents three main problems. First, it works well with convex budget sets (e.g. those generated by progressive taxation) and a two-good application (e.g. leisure and income in the individual labour supply model) but it tends to become computationally cumbersome when the decision makers face non-convex budget sets and when more than two goods are object to choice (e.g. in the case of a many-person household). Second, in view of the computational problems, the above approach essentially forces the researcher to choose relatively simple specifications for the utility function or the labour supply functions. Third, computational and statistical consistency of ML estimation of the model requires imposing a priori quasi-concavity of the utility function (e.g. see MaCurdy et al., 1990).

As a response to the problems mentioned above, researchers have since the early 80s made use of another innovative research effort which matured in the first half of the 70's, i.e. the random utility maximization (RUM) model developed by McFadden (1974, 1981). It is not often realized in the literature that the advantages of this approach (as we will explain more precisely in section 2.1) are due to the representation of choice as the maximization of a random utility, rather than to the discreteness of the choice set. In practice, however, the most common implementation of the approach involves a discrete representation of the choice set. As far as the labour supply application is concerned, this approach essentially consists in representing the budget set with a set of discrete alternatives or jobs. The choice of the optimal alternative is modelled in terms of a comparison between utility level and not in terms of conditions involving marginal utilities. Allowing the utility function to be stochastic and using a convenient specification for the stochastic component (i.e. the extreme value distribution) leads to an easy and intuitive expression for the probability that any particular point is chosen (i.e. the Multi-Nomial Logit model). This approach is very convenient when compared to the previous ones, since it does not require going through complicated Kuhn-Tucker conditions involving derivatives of the utility function and of the budget constraints. As a consequence it is not affected by the complexity of the rule that defines the budget set or by how many goods are contained in the utility function. Equally important, the deterministic part of the utility function can be specified in a very flexible way without worrying about the computational problems.

During the last two decades, this approach has become increasingly popular in the labour supply literature, starting with Aaberge et al (1995) and van Soest (1995). Within the literature adopting this approach there are however two issues which have not been given the attention we think they deserve.

A first issue concerns the procedure by which the discrete alternatives are included in the choice set. Most authors (e.g., among others, van Soest (1995), Duncan and Weeks (1997), Blundell, Duncan et al. (2000)), Kornstad and Thoresen (2004)) choose (not probabilistically) a set of fixed points which is identical for each individual[1]. By contrast, Aaberge et al. (1995) and Aaberge et al. (1999) adopt a sampling procedure originally proposed by McFadden (1978).

A second issue concerns the availability of the alternatives. Letting H represent the maximum number of hours in the reference period, most authors assume all the values in $[0, H]$ - or in some discrete subset - are equally available. At the other extreme, some authors (e.g. Zabalza et al. (1980) assume only two or three alternatives (e.g. non-participation, part-time and full-time) are available for everyone. Aaberge et al. (1995, 1999, 2000, 2004) assume instead that all the hour opportunities in $[0, H]$ are in principle available but not equally accessible for everyone. More specifically, they assume that there is a probability density function of opportunities for each individual. The discrete choice set used in the estimation (and subsequently in the simulations) is built by sampling from that individual-specific density function.

Section 2 explains in more detail the implications of alternative procedures used to generate the choice set and defines the different types of models that can be estimated accordingly. Sections 3, 4 and 5 present the simulation exercises. We use a previously estimated model of female labour supply as the "true" model. The model (described in Section 3) is characterized by heterogeneous availability of alternatives (across different hour values and among different individuals). From the "population" described by the "true" model we generate 30 samples for a Monte Carlo exercise. In Section 4 we use the data from these samples to estimate – and compare the prediction performance of – various models that adopt the same specification of preferences as in the "true" model but differ in the way the choice set is represented (sampled vs. fixed alternatives, number of alternatives, heterogeneous vs. uniform availability of alternatives). In Section 5 we perform a second simulation exercise where we focus more deeply on the systematic impact of different specifications of the choice set upon the in-sample and out-of-sample prediction error. Section 6 contains the conclusions.

2. Alternative representations of the choice sets

In this section, after recalling the basic discrete choice version of the labour supply model, we survey the crucial problems to be faced in specifying the choice set, i.e. the selection of the alternatives and the representation of different availability of alternatives.

2.1 The basic Random Utility Maximization (RUM) model of labour supply

The individuals maximize their utility by choosing from opportunities ("jobs") defined by hours of work and other unobserved (by the analyst) attributes. The utility is assumed to be of the following form

(2.1) $$U\big(f(wh,I),h,j\big)=v(f(wh,I),h)+\varepsilon(j)$$

where w is the wage rate, h is hours of work, I is exogenous income, f is a tax-transfer function that transforms gross incomes into net income, j is a variable that captures other job and/or individual characteristics and ε is a random variable. Commuting time or required skill are possible examples of the characteristics captured by j. The model as specified in (2.1) belongs to the class of the Random Utility Maximization (RUM) models (see for example McFadden 1981). Let $B=\begin{bmatrix}0,H\end{bmatrix}$ be the range of possible values for hours of work h and let $p(h)$ be the probability density function of jobs with hours equal to h. The most common distribution to assume for the random term ε is the Type I Extreme Value[2]. If the range of values of h is continuous, the stochastic assumption leads to the (continuous) multinomial logit expression for the probability that a job with h hours is chosen[1]:

(2.2) $$\varphi(h)\equiv\Pr\left[U(f(wh,I),h)=\max_{x\in B}U(f(wx,I),x)\right]=\frac{\exp(v(f(wh,I),h)\,p(h)}{\int\exp(v(f(wx,I),x)\,p(x)dx}.$$

Based on (2.2), the corresponding likelihood function can then be computed and maximized in order to estimate the parameters of the utility function. The crucial advantage of this approach is that the characterization of the utility maximization problem (i.e. expression (2.1)) is not affected by the specification of v nor of f. In other words, one can choose relatively general and complicated specifications for v and/or accounting for complex tax-transfer rules f without affecting the characterization of behaviour and without significantly affect the computational burden involved by the estimation or simulation of the model. Expression (2.2) is a simplified version the model developed by Dagsvik (1994) and by Aaberge et al. (1999). It is also close to the continuous spatial model developed by Ben-Akiva and Watanatada (1981). We have chosen to start with the continuous version of the multinomial logit model in order to highlight the fact that the advantages of the approach are due not so much to a discrete representation of the choice set but rather to the specification of utility as a random variable. Although in principle the model could be directly managed in the form expressed by (2.2), in practice, for ease of interpretation, a discrete representation is usually preferred. Clearly the researcher might think that the choice set, at least as it is perceived by the household, is in essence discrete; but even a genuinely continuous range of values can always be represented (to any desirable degree of approximation) by a set of discrete values. The probability that a job with hours equal to h is chosen can therefore be written as follows:

---

[1] Note that Aaberge et al. (1995, 1999, 2000, 2004) consider B to be the set of market as well non-market opportunities where market opportunities (jobs) are characterized by hours of work as well as by the wage rate and other job attributes.

$$(2.3) \qquad \varphi(h) = \frac{\exp(v(f(wh,I),h)\,p(h)}{\sum\limits_{x \in B} \exp(v(f(wx,I),x)\,p(x)}.$$

A further common simplification (mostly implicit in the literature on labour supply) is assuming that all the values in B are equally frequent (or dense), i.e. $p(h) = a$ (constant) *for all h*. With this assumption we get

$$(2.4) \qquad \varphi(h) = \frac{\exp(v(f(wh,I),h)}{\sum\limits_{x \in B} \exp(v(f(wx,I),x)}.$$

2.2 Selection of alternatives

As we have already mentioned in Section 1, the first issue in choice set representation concerns the procedure used to select the alternatives. In many applications, including labour supply modelling, the choice set contains a very large (or even infinite) number of alternatives. For instance, if we model labour supply of couples and the decision period is the year, considering 1 hour intervals and 16 hours available during the day, there are $(16 \times 365)^2 = 34,105,600$ alternatives. This would imply a very heavy computational burden, since for each alternative we must compute the couple's budget by applying a possibly complicated tax rule. More in general, if the alternatives are characterized by K attributes and the k-th attribute can take Q_k different values, the choice set contains

$\prod_{k=1}^{K} Q_k$ alternatives. Thus it is convenient to work with a smaller choice set somehow representative of the true one. Ben-Akiva and Lerman (1985) present a detailed treatment based on either aggregating alternatives or sampling alternatives when the number of alternatives contained in the choice set is very large (or even infinite) so that a complete enumeration is computationally too costly. For the sake of simplicity, we will in this section refer to the representation expressed by (2.4), where the assumption is that all the alternative values of h are equally available (i.e. equally frequent in the choice set). The issue of a non-uniform availability of alternatives will be addressed in Section 2.3.

Aggregating alternatives.

The procedure consisting in selecting a fixed number of hours values can be interpreted as an aggregation procedure. Instead of using all the possible values between 0 and H, the [0, H] range is divided into sub-intervals and then the mid (or maybe the average) value of h in each interval is chosen to 'represent' all the values of that interval. The authors adopting this procedure realize that it introduces measurement errors, but tend to assume they are of minor importance. For example van Soest (1995) reports that some experiments with a different number of points did not show significant differences in parameter estimates. However a systematic investigation of the implication of that procedure has never been done either theoretically or empirically.

If one interprets the approximation of the choice sets as an aggregation procedure, the analysis provided by Ben-Akiva and Lerman (1985) can be applied to clarify the issue. The interval [0, H] is divided into L sub-intervals. We will assume the average of h in each sub-interval is chosen as representative (instead of the more common procedure of choosing the mid-point: of course the two are very close and in fact coincide if the values of h are continuous or if each interval contains an uneven number of values). Using the terminology introduced in Section 2.1, let

$$\overline{v}^{\ell} \equiv \frac{1}{N^{\ell}} \sum_{h \in B_{\ell}} v\big(f(wh, I), h\big) = \text{average systematic utility in sub-interval } \ell, \text{ where } B_{\ell} \text{ is the set of values}$$

of hours contained in sub-interval ℓ and N^{ℓ} is the number of elements contained in B_{ℓ}. Ben-Akiva and Lerman (1985) show that the expected maximum utility attained on subinterval ℓ is

$$(2.5) \qquad \qquad \widehat{v}^{\ell} = \overline{v}^{\ell} + \ln(N^{\ell}) + \ln\big(D^{\ell}\big)$$

where $D^{\ell} \equiv \sum_{j} \exp\big(v_{j}^{\ell} - \overline{v}^{\ell}\big) \frac{1}{N^{\ell}}$. This last term is a measure of dispersion of v in sub-interval ℓ.

Accordingly, the probability that a value of h belonging to sub-interval ℓ is chosen is

$$(2.6) \qquad \qquad \varphi(\ell) = \frac{\exp\big(\overline{v}^{\ell} + \ln(N^{\ell}) + \ln\big(D^{\ell}\big)\big)}{\displaystyle\sum_{i=1}^{L} \exp\big(\overline{v}^{i} + \ln(N^{i}) + \ln\big(D^{i}\big)\big)} .$$

To compare this with the expression used in the fixed-alternatives approach it is useful to Taylor-expand v_{j} up to 2-order terms to get

$$(2.7) \qquad \qquad \varphi(\ell) \cong \frac{\exp\big(v\big(f(w\overline{h}^{\ell}, I), \overline{h}^{\ell}\big) + 0.5\sigma_{hh}^{\ell} v_{hh}^{\ell} + \ln(N^{\ell}) + \ln\big(D^{\ell}\big)\big)}{\displaystyle\sum_{i=1}^{L} \exp\big(v\big(f(w\overline{h}^{i}, I), \overline{h}^{i}\big) + 0.5\sigma_{hh}^{i} v_{hh}^{i} + \ln(N^{i}) + \ln\big(D^{i}\big)\big)}$$

where $\overline{h}^{i}$ is the average of h in sub-interval i, σ_{hh}^{i} is the variance of h in sub-interval i and v_{hh}^{i} is the second (total) derivative of $v\big(f(w\overline{h}^{i}, I), \overline{h}^{i}\big)$ evaluated at $h = \overline{h}^{i}$. It would be pointless to use expression (2.7) for estimation since it requires the very same computations that one wishes to avoid by aggregating alternatives. However expression (2.7) is useful in order to understand the type and the extent of the errors we incur by using various approximations. The expression typically used in the literature is:

$$(2.8) \qquad \qquad \varphi(\ell) \cong \frac{\exp\big(v\big(f(w\overline{h}^{\ell}, I), \overline{h}^{\ell}\big)\big)}{\displaystyle\sum_{i=1}^{L} \exp\big(v\big(f(w\overline{h}^{i}, I), \overline{h}^{i}\big)\big)} .$$

8

In expression (2.8) all the terms $0.5\sigma_{hh}^{\ell}v_{hh}^{\ell} + \ln(N^{\ell}) + \ln\left(D^{\ell}\right)$ appearing in (2.7) are dropped. If these terms were equal across all the sub-intervals they would cancel out from (2.7) and (2.8) would be exact. In general however they will not be equal, and dropping them will lead to biased estimates. Nonetheless there are ways by which we could improve upon (2.8) when adopting aggregation as an approximation strategy; ways which however has never been considered in the literature on labour supply modelling:

- The dimension of N^{i} of the sub-intervals - when not equal for all of them - is typically known and can be explicitly accounted for;

- σ_{hh}^{i} can also be computed;

- Depending on the functional form used for the utility function, the term v_{hh}^{i} might be explicitly evaluated and accounted for;

- The terms $\ln\left(D^{i}\right)$ in general will vary both across sub-intervals and across individuals; however we might capture at least some of their effects by introducing a set of dummies (as many as the number of sub-intervals - 1).

Summing up, the aggregation of alternatives implies biased estimates. The bias could be moderated by using various possible corrections suggested by expression (2.7). However, it must be said that the literature on labour supply so far has treated this issue in a rather superficial way (as compared, for instance, to the literature on transportation or on location choices).

Sampling alternatives.

Sampling of alternatives, on the other hand, offers the possibility of working with a relatively small choice set and at the same time preserving the consistency of the estimates. The basic results are established by McFadden (1978). Ben-Akiva and Lerman (1985) also provide a very useful and more practically oriented survey, together with some additional theoretical results. Let us represent the true choice set B with a sample S containing a subset of the alternatives contained in B, where one alternative is the chosen (observed) point and the others are sampled from a probability density function *q(h)*. It can be shown (McFadden, 1978; Ben Akiva and Lerman, 1985) that consistent estimates of $v(f(wh,I),h)$ can still be obtained when the true choice set B is replaced by S and the probability of observing choice *h* is evaluated as follows:

(2.9)
$$\varphi(h|S) = \frac{\exp\left(v(f(wh,I),h) - \ln(q(h))\right)}{\sum_{x \in S}\exp\left(v(f(wx,I),x) - \ln(q(x))\right)}.$$

If a simple random sampling is adopted, all the *q's* are equal and cancel out. Typically more sophisticated sampling procedures are used since they are expected to be more efficient. For instance, a common procedure consists of using as sampling probabilities the observed relative frequencies of choice possibly differentiated according to personal characteristics of the decision units. Besides Ben-Akiva and Lerman (1985), also Train et al. (1987) and Colombino (1998) present a very detailed application of this procedure.

2.3 Availability of alternatives

A second and possibly even more substantial issue is whether account is taken of the different availability of job-types on the market. Some authors have made the extreme choice of assuming that the choice set contains only two or three alternatives (e.g. non-participation, part-time and full-time). More common, however, is the approach of choosing a few equally spaced points in the interval [0,H], without taking into account the possibility that some type of opportunities might be more easily available than others. Other authors (Aaberge et al. 1995, 1999, 2004) do account for this possibility as well as for the relative density of jobs as a function of personal characteristics. This implies using (2.3) instead of (2.4) as the choice probability. In practice, based on a convenient specification of the probability density function *p(h)* the procedure boils down to "augmenting" the term *v* with a set of appropriately defined dummy variables. Van Soest (1995) introduces similar dummies and interprets them as reflecting costs or benefits and search costs attached to specific ranges of hours values.[3]

3. The simulation exercise

In the following sections we illustrate the results of two simulation exercises. The first one is a Monte Carlo simulation and consists of three steps. First, we use a previously estimated model of married women's labour supply (the "true" model illustrated in Section 3.1) to draw 30 samples; each with 1842 observations. In other words, the parameters of the "true" model are treated as the population parameters. The samples are generated by drawing 30 values of the random component (Type I – extreme value distributed) of the utility function for each individual in the original estimation sample (1842 observations). Correspondingly we compute 30 optimal choices for each individual. As a result we obtain 30 samples of 1842 observations. Second, various specific models adopting different representations of the choice set (the details are given in Section 3.2) are estimated on the 30 samples. Thus, for each type of model we obtain a set of 30 estimates. Third, we evaluate the performance of the different models by comparing the models' predictions with the values – as predicted by the "true" model – of income, participation and hours of work. The evaluation of the prediction performance is made in-sample as well as out-of-sample. The in-sample evaluation consists in comparing the values predicted by the "true" model to the values predicted by each alternative model. In the out-of-sample exercise we first use the "true" model to simulate the effects of a tax reform (a revenue-constant flat

tax); next, we compare the simulated "true" values to those obtained by simulating the various alternative models under the same tax reform. We report the mean and the standard deviation (computed on the 30-sample distribution) of the prediction errors.

Since it turns out that the performance of the models differs only in the mean of the prediction error but not in the standard deviation of the prediction error, in the second simulation exercise we focus on the mean prediction error and on its relationship with the characteristics of the different alternative models. In this second exercise we simulate the drawing of a large sample from the population (again defined by the parameters of the "true" model). We use a large sample in order to minimize the "noise" due to sampling variations and focus on the systematic differences between the models. The sample is formed by drawing 6 values of the random component (Type I – extreme value distributed) of the utility function for each individual in the original estimation sample (1842 observations). Correspondingly we compute 6 optimal choices for each individual. As a result we get a large sample of $6 \times 1842 = 11052$ observations. The different types of models are then estimated on this large sample. For each model we compute an index of prediction performance and then regress the index on a set of variables measuring the different characteristics of the model in order to identify the contribution of the different characteristics to the prediction performance.

3.1. The "true" model

The "true" model is defined as in expressions (2.1) and (2.2) and empirically specified along the lines adopted in Aaberge et al. (1995) as well as in several successive papers.[4] We model the choice of married/cohabitating females, and maintain other household members' behaviour as exogenous. The systematic part of the utility function is specified as follows

(3.1)

$$v(f(wh,I),h) = \alpha_2 \left(\frac{f(wh,I)^{\alpha_1} - 1}{\alpha_1} \right) + (\alpha_4 + $$
$$+ \alpha_5 \log A + \alpha_6 (\log A)^2 + \alpha_7 C_1 + \alpha_8 C_2 + \alpha_9 C_3) \left(\frac{L^{\alpha_3} - 1}{\alpha_3} \right)$$

where L is a measure of leisure, defined as $L = 1 - (h/8736)$ and h is yearly hours of work, A is age and C_1, C_2 and C_3 are number of children below 3, between 3 and 6 and between 7 and 14 years old.

We specify the density of opportunities requiring h hours of work as

$$(3.2) \qquad p(h) = \begin{cases} p_0 g(h) & \text{if} \quad h > 0 \\ 1 - p_0 & \text{if} \quad h = 0 \end{cases}$$

where p_0 is the proportion of market opportunities in the opportunity set, and g is the density of hours conditional upon the opportunity being a market job (i.e. $h > 0$).

Offered hours are assumed to be uniformly distributed except for possible peaks at half-time (corresponding to 18-20 weekly hours), and to full-time (corresponding 37-40 weekly hours). Thus, g is given by

$$(3.3) \qquad g(h) = \begin{cases} \gamma & \text{if} \quad h \in (52, 910] \\ \gamma \exp(\pi_1) & \text{if} \quad h \in (910, 1066] \\ \gamma & \text{if} \quad h \in (1066, 1898] \\ \gamma e \operatorname{xp}(\pi_2) & \text{if} \quad h \in (1898, 2106] \\ \gamma & \text{if} \quad h \in (2106, 3640] \end{cases}$$

where H is the maximum observed value of h. Thus, this opportunity density for offered hours implies that it is more likely to find jobs with hours that accord with full-time and standard part time positions than jobs with other working loads.[5]

Based on (3.2) and (3.1) and using the definitions

$$(3.4) \qquad \frac{p_0}{1 - p_0} = \exp(\theta_0)$$

$$d_0(h) = 1 \text{ if } h > 0; \ 0 \text{ otherwise}$$
$$d_1(h) = 1 \text{ if } h \in [910, 1066]; \ 0 \text{ otherwise}$$
$$d_2(h) = 1 \text{ if } h \in [1898, 2106]; \ 0 \text{ otherwise}$$

the probability that an opportunity with h hours of work is chosen (i.e. expression (2.2) can be rewritten as follows:

$$(3.5) \qquad \varphi(h) = \frac{\exp\big(v\big(f(wh, I), h\big) + \theta_0 d_0(h) + \pi_1 d_1(h) + \pi_2 d_2(h)\big)}{\int \exp\big(v\big(f(wx, I), x\big) + \theta_0 d_0(x) + \pi_1 d_1(x) + \pi_2 d_2(x)\big) dx}.$$

.

We refer to π_1, π_2 and θ_0 as the parameters of the *opportunity density*. In what follows we will refer to d_0 as the "job" dummy, since it captures the relative frequency of market opportunities to non-market opportunities; we will refer to d_1 and d_2 as the "peaks" dummies, since they are meant to capture the "peaks" in the density of hours corresponding to part-time and full-time jobs.

The parameters of the utility function (3.1) and the parameters of the job opportunity density defined by (3.2) and (3.3) are estimated by maximum likelihood. The continuous choice set is approximated by a discrete choice set S containing the chosen value of h plus 999 values sampled from the empirical probability density function $q(h)$. Then, using one of the procedures explained in McFadden (1978) and Ben Akiva and Lerman (1985), consistent estimates of the parameters can be obtained by using the following expression for the individual contribution to the likelihood function:

$$(3.6) \quad \varphi(h|S) = \frac{\exp\left(v\left(f(wh,I),h\right) + \theta_0 d_0(h) + \pi_1 d_1(h) + \pi_2 d_2(h) - \ln(q(h))\right)}{\sum_{x \in S} \exp\left(v\left(f(wx,I),x\right) + \theta_0 d_0(x) + \pi_1 d_1(x) + \pi_2 d_2(x) - \ln(q(x))\right)}.$$

The estimation of the model is based on data for 1842 married/cohabitating females from the 1995 Norwegian Survey of Level of Living. We have restricted the ages of the females to be between 20 and 62 years in order to minimize the inclusion in the sample of individuals who in principle are eligible for retirement, since analysis of retirement decisions is beyond the scope of this study. Although the model adopted was originally developed for analysing simultaneous household partners' behaviour, we focus here on women's behaviour in order to simplify the execution and the interpretation of the simulation exercise. Moreover, the majority of labour supply studies have primarily focused on married/cohabitating females, where husband's income as well as the couple's non-labour income are treated as exogenous and included in disposable income $f(wh,I)$.[6] The estimates are presented in Table A.1 of Appendix A.

3.2. Alternative models

In what follows we use the sample generated according to the true model to estimate various versions of models generated according to the various possible representations of the choice set as discussed in Section 2.

The more general versions of the models are (3.6) when sampled alternatives are used, and

$$(3.7) \quad \varphi(h|R) = \frac{\exp\left(v\left(f(wh,I),h\right) + \theta_0 d_0(h) + \pi_1 d_1(h) + \pi_2 d_2(h)\right)}{\sum_{x \in R} \exp\left(v\left(f(wx,I),x\right) + \theta_0 d_0(x) + \pi_1 d_1(x) + \pi_2 d_2(x)\right)}$$

when fixed alternatives are used. R denotes the choice set built as a set of fixed alternatives. The dummies d_0 and (d_1, d_2) are defined as in (3.4). Dropping the job dummy d_0 and/or the peaks dummies (d_1, d_2) generates a more restrictive version of the model. The choice sets S and R contain alternatively 6 or 24 points. For the model with fixed alternatives, we choose the mid-values of (6 or

24) equally spaced intervals between 0 and 3640. For the model with sampled alternatives, the choice set contains the observed value of h plus 5 or 23 values sampled from the empirical distribution g (defined by (3.3)) of offered hours.

Altogether we have 16 models resulting from the combinations of the following possibilities:

1. *alternative generation*: fixed or sampled;
2. *number of alternatives*: 6 or 24;
3. *job dummy*: included or dropped;
4. *peaks dummies*: included or dropped.

The Tables that report the results of the 16 models are labelled as in Table 3.1. The parameter estimates of the 16 models are reported in the Appendix (Tables A.2). [7]

We are interested in the prediction performance of the models, both in-sample and out-of-sample (prediction of policy effects). Clearly, we expect the more flexible and complex models (i.e. those allowing for a different availability of alternatives) to perform better than simpler or more restrictive models. Also, we know that the models based on sampled alternatives are expected to produce consistent estimates, while those based on fixed alternatives are not. Therefore what in fact we want to explore is *how much better* the more flexible models perform and *how much better* the models based on sampled alternatives perform.

Table 3.1. Types of models

	Generation of alternative	*Number of alternatives*	*Job dummy*	*Peaks dummies*
Model Ia	Fixed	6	No	No
Model Ib	Fixed	6	Yes	No
Model Ic	Fixed	6	No	Yes
Model Id	Fixed	6	Yes	Yes
Model IIa	Fixed	24	No	No
Model IIb	Fixed	24	Yes	No
Model IIc	Fixed	24	No	Yes
Model IId	Fixed	24	Yes	Yes
Model IIIa	Sampled	6	No	No
Model IIIb	Sampled	6	Yes	No
Model IIIc	Sampled	6	No	Yes
Model IIId	Sampled	6	Yes	Yes
Model IVa	Sampled	24	No	No
Model IVb	Sampled	24	Yes	No
Model IVc	Sampled	24	No	Yes
Model IVd	Sampled	24	Yes	Yes

4. A Monte Carlo exercise

In this exercise, each model is estimated on the 30 samples obtained as explained in Section 3. For each model and each of the 30 repetitions we predict participation rates, hours of work and disposable income. The predictions are obtained individual by individual, evaluating the utility function – including the random component drawn from the Type I extreme value distribution – at each alternative and identifying the selected alternative as the one with the highest utility level. The individual predictions are then aggregated into the 10 means of the 10 income deciles. We define the relative prediction error as follows:

4.1
$$z_{kjs} = \frac{\tilde{y}_{kjs} - y_j}{y_j} \, , \, j = 1,..., 10; \, k=1, ...,4; \, s = 1,..., 30;$$

where y_j and $\tilde{y}_{kjs}$ denote the outcomes in decile j of the true model and alternative model k in sample s, respectively. The outcomes are alternatively defined to be the job participation rate, hours of work and disposable income after tax. The exercise is done twice, once for predicting the current (1994) values (and comparing them with those predicted by the "true" model) and once for predicting the effects of a hypothetical revenue-constant Flat Tax (and comparing them with those predicted by the "true" model).

In order to simplify the presentation Tables 4.1 – 4.6 report the results only for the four models Ia, IIb, IIIc and IVd.[8] The left part of each table contains the means of the relative prediction error, i.e.

$\bar{z}_{kj} = \sum_{s=1}^{30} z_{kjs} / 30$, while the right part contains the standard deviations, i.e. $\sqrt{\sum_{s=1}^{30} \left(z_{kjs} - \bar{z}_{kj} \right)^2 / 30}$.

From the tables we can observe that

1) Sampled alternative models (IIIc and IVd) perform better than fixed alternatives models (Ia and IIb.

2) Predictions tend to be less precise in lower and upper deciles, more notably so with model Ia. This result is in accord with what one would expect because a simplification of a model normally is not costless. A poorer description of the choice set weakens the model's ability to predict the tails of the distributions.

3) There are no notable differences in the standard deviation of prediction error among the models.

Table 4.1. Mean and standard deviation of the relative differences between disposable income in the true model and 4 different models under the 1994 tax system

Mean					Std.dev.			
Model Ia	Model IIb	Model IIIc	Model IVd	Income decile	Model Ia	Model IIb	Model IIIc	Model IVd
0.9 %	1.2 %	1.5 %	1.2 %	1	1.3 %	1.3 %	1.2 %	1.2 %
-0.4 %	-0.4 %	-0.5 %	-0.6 %	2	0.9 %	1.0 %	0.8 %	0.9 %
-0.7 %	-0.9 %	-1.2 %	-1.1 %	3	0.6 %	0.8 %	0.7 %	0.8 %
0.3 %	0.2 %	0.0 %	0.2 %	4	0.8 %	0.7 %	0.6 %	0.6 %
0.7 %	0.5 %	0.3 %	0.6 %	5	0.8 %	0.7 %	0.5 %	0.5 %
0.1 %	0.0 %	-0.2 %	0.1 %	6	0.7 %	0.6 %	0.5 %	0.5 %
-0.4 %	-0.5 %	-0.7 %	-0.4 %	7	0.7 %	0.6 %	0.5 %	0.4 %
-0.4 %	-0.7 %	-0.7 %	-0.5 %	8	0.5 %	0.5 %	0.5 %	0.4 %
-0.1 %	-0.7 %	-0.4 %	-0.4 %	9	0.5 %	0.5 %	0.6 %	0.6 %
2.0 %	0.8 %	0.9 %	0.8 %	10	0.6 %	0.5 %	0.6 %	0.6 %
0.3 %	0.0 %	-0.1 %	0.0 %	All	0.3 %	0.4 %	0.3 %	0.3 %

Table 4.2. Mean and standard deviation of the relative differences between participation rate in the true model and 4 different models under the 1994 tax system

Mean					Std.dev.			
Model Ia	Model IIb	Model IIIc	Model IVd	Income decile	Model Ia	Model IIb	Model IIIc	Model IVd
-7.7 %	0.5 %	19,9 %	3,5 %	1	6,3 %	4,6 %	4,6 %	4,7 %
5,0 %	4,6 %	17,8 %	5,2 %	2	6,4 %	6,4 %	6,4 %	6,7 %
-0,3 %	-3,6 %	3,1 %	-3,1 %	3	3,5 %	3,8 %	3,2 %	3,3 %
2,2 %	-1,0 %	2,4 %	-1,6 %	4	2,9 %	2,9 %	3,1 %	3,3 %
-1,3 %	-2,2 %	-0,1 %	-2,0 %	5	2,2 %	1,8 %	2,1 %	2,4 %
1,5 %	-0,1 %	1,8 %	0,2 %	6	1,4 %	1,9 %	1,6 %	1,6 %
1,2 %	0,0 %	2,1 %	1,0 %	7	1,4 %	1,7 %	1,3 %	1,3 %
-0,5 %	-2,1 %	-0,8 %	-2,4 %	8	1,4 %	1,5 %	2,2 %	2,2 %
0,4 %	-0,7 %	0,6 %	-0,4 %	9	1,5 %	1,3 %	0,9 %	1,0 %
5,7 %	0,9 %	5,0 %	2,4 %	10	2,3 %	2,0 %	2,7 %	2,5 %
0,8 %	-0,5 %	4,1 %	0,0 %	All	1,0 %	0,9 %	0,9 %	0,9 %

Table 4.3. Mean and standard deviation of the relative differences between hours of work in the true model and 4 different models under the 1994 tax system

Mean					Std.dev.			
Model Ia	Model IIb	Model IIIc	Model IVd	Income decile	Model Ia	Model IIb	Model IIIc	Model IVd
0.0 %	0.0 %	0.0 %	0.0 %	1	0.0 %	0.0 %	0.0 %	0.0 %
7.6 %	0.1 %	-0.7 %	-3.4 %	2	8.0 %	6.6 %	7.4 %	7.8 %
4.0 %	-2.7 %	-5.4 %	-5.1 %	3	6.0 %	6.1 %	6.2 %	6.8 %
0.6 %	-2.1 %	-4.1 %	-3.4 %	4	5.0 %	3.9 %	5.9 %	5.1 %
2.4 %	1.2 %	2.2 %	4.9 %	5	4.3 %	4.0 %	3.5 %	3.9 %
-1.1 %	-3.5 %	-3.9 %	-2.1 %	6	2.9 %	3.3 %	3.4 %	3.5 %
2..6 %	0..3 %	1..1 %	2..1 %	7	3..2 %	3..1 %	2..8 %	3..0 %
1..6 %	-1.8 %	-2.2 %	-1.7 %	8	2.7 %	2.9 %	3.3 %	3.4 %
3.0 %	-1.0 %	-1.9 %	-1.0 %	9	2.3 %	2.7 %	2.9 %	2.9 %
11.3 %	3.3 %	6.3 %	5.5 %	10	3.1 %	3.0 %	3.4 %	3.5 %
3.7 %	-0.2 %	0.0 %	0.3 %	All	1.3 %	1.5 %	1.2 %	1.2 %

Table 4.4. Mean and standard deviation of the relative differences between disposable income in the true model and 4 different models under a flat tax reform

Mean					Std.dev.			
Model Ia	Model IIb	Model IIIc	Model IVd	Income decile	Model Ia	Model IIb	Model IIIc	Model IVd
-13.2 %	-8.4 %	-8.8 %	-9.0 %	1	1.9 %	2.0 %	1.9 %	1.9 %
-12.2 %	-8.3 %	-7.2 %	-7.9 %	2	1.5 %	1.6 %	1.8 %	1.6 %
-7.0 %	-3.9 %	-4.4 %	-4.6 %	3	1.3 %	1.6 %	1.5 %	1.3 %
-6.8 %	-4.4 %	-4.5 %	-4.7 %	4	1.1 %	1.0 %	1.3 %	1.2 %
-4.3 %	-1.8 %	-2.2 %	-2.4 %	5	0.8 %	0.8 %	0.8 %	1.0 %
-4.9 %	-2.9 %	-2.4 %	-2.5 %	6	0.8 %	0.7 %	0.9 %	0.9 %
-2.0 %	-0.3 %	-0.4 %	-0.4 %	7	0.8 %	1.0 %	1.0 %	1.0 %
-4.3 %	-3.1 %	-3.1 %	-3.2 %	8	0.8 %	0.7 %	1.0 %	0.7 %
-2.2 %	-1.2 %	-0.8 %	-1.0 %	9	0.8 %	0.9 %	0.9 %	0.9 %
0.9 %	0.6 %	1.0 %	0.9 %	10	0.6 %	0.6 %	0.7 %	0.8 %
-4.3 %	-2.5 %	-2.4 %	-2.6 %	All	0.3 %	0.4 %	0.3 %	0.4 %

Table 4.5. Mean and standard deviation of the relative differences between participation rate in the true model and 4 different models under a flat tax reform

Mean					Std.dev.			
Model Ia	Model IIb	Model IIIc	Model IVd	Income decile	Model Ia	Model IIb	Model IIIc	Model IVd
-14.1 %	-3.7 %	9.4 %	-1.5 %	1	5.3 %	4.4 %	3.6 %	4.0 %
-6.7 %	-1.8 %	8.1 %	-1.4 %	2	5.7 %	5.5 %	3.8 %	5.1 %
-1.5 %	-1.9 %	3.3 %	-1.6 %	3	3.4 %	3.6 %	3.1 %	3.2 %
-0.6 %	-1.8 %	1.4 %	-2.3 %	4	2.7 %	2.7 %	2.7 %	3.2 %
-1.8 %	-1.5 %	0.1 %	-1.9 %	5	2.4 %	1.7 %	2.0 %	2.1 %
-0.2 %	-0.9 %	0.5 %	-0.9 %	6	1.3 %	1.5 %	1.4 %	1.6 %
-0.1 %	-0.9 %	1.4 %	0.2 %	7	1.5 %	1.7 %	1.5 %	1.6 %
-0.2 %	-1.5 %	0.1 %	-1.3 %	8	1.3 %	1.5 %	1.9 %	2.1 %
0.5 %	-0.3 %	1.0 %	0.2 %	9	1.3 %	1.1 %	1.0 %	0.9 %
4.9 %	1.0 %	4.6 %	2.4 %	10	2.1 %	2.0 %	2.6 %	2.5 %
-1.5 %	-1.2 %	2.6 %	-0.8 %	All	1.0 %	0.9 %	0.7 %	0.9 %

Table 4.6. Mean and standard deviation of the relative differences between hours of work in the true model and 4 different models under a flat tax reform

Mean					Std.dev.			
Model Ia	Model IIb	Model IIIc	Model IVd	Income decile	Model Ia	Model IIb	Model IIIc	Model IVd
-18.3 %	-8.2 %	-5.3 %	-8.5 %	1	0.6 %	13.3 %	15.7 %	10.7 %
-21.9 %	-15.4 %	-13.5 %	-16.6 %	2	6.3 %	5.4 %	5.4 %	6.7 %
-6.5 %	-2.6 %	-5.6 %	-5.6 %	3	5.8 %	4.9 %	4.5 %	6.1 %
-9.7 %	-6.7 %	-7.6 %	-8.3 %	4	4.4 %	3.9 %	5.5 %	4.9 %
-3.4 %	0.9 %	1.5 %	2.8 %	5	4.1 %	2.9 %	3.7 %	3.9 %
-6.2 %	-5.1 %	-4.8 %	-4.2 %	6	2.6 %	2.5 %	3.2 %	3.1 %
1.9 %	2.1 %	3.5 %	4.0 %	7	3.4 %	3.0 %	3.0 %	2.7 %
-0.6 %	-1.6 %	-1.6 %	-1.2 %	8	2.5 %	2.9 %	3.6 %	3.3 %
2.8 %	1.2 %	0.8 %	1.2 %	9	2.2 %	2.8 %	2.7 %	2.5 %
10.6 %	4.8 %	8.3 %	7.7 %	10	2.9 %	3.2 %	3.3 %	3.5 %
-3.6 %	-2.3 %	-1.7 %	-1.9 %	All	1.1 %	1.1 %	1.0 %	1.2 %

5. Choice set representation and prediction performance: a systematic analysis.

In this section we evaluate the impact of alternative representations of the choice set on the performance of the models. As explained in Section 3, we use the large sample of $1842 \times 6 = 11052$ observations in order to neglect the effect of sampling variations and focus on the systematic differences among alternative representations of the choice set. First, for each of the 16 models (see Table 3.1) we predict participation rates, hours of work and disposable income. As with the previous

exercise illustrated in Section 4, the predictions are obtained individual by individual, by evaluating the utility function – including the stochastic component drawn from the Type I extreme value distribution – at each alternative and identifying the selected alternative as the one with the highest utility level. The individual predictions are then aggregated into the 10 means of the 10 income deciles. We introduce the following summary measure of prediction performance (relative prediction error) z_k for model k,

$$(5.1) \qquad z_k = \left| \sqrt{\sum_{j=1}^{10} \left(\frac{(\tilde{y}_{kj} - y_j)}{y_j} \right)^2} \right|, \; k=1, 2 \ldots,16,$$

where y_j and $\tilde{y}_{kj}$ denote the outcomes in decile j of the true model and alternative model k, respectively. The outcomes are alternatively defined to be the job participation rate, hours of work and disposable income after tax. We define:

$x_{1k}= 1$ if the choice alternatives are sampled (= 0 if the choice alternatives are fixed),

$x_{2k}= 1$ if the number of choice alternatives is equal to 24 (= 0 if the number of alternatives is equal to 6),

$x_{3k} = 1$ when a job dummy is included (= 0 otherwise),

$x_{4k} = 1$ when peaks dummies are included (= 0 otherwise).

We then estimate the following regression equation[9]

$$(5.2) \qquad \begin{aligned} \ln(z_k) = &\alpha_0 + \alpha_1 x_{1k} + \alpha_2 x_{2k} + \alpha_3 x_{3k} + \alpha_4 x_{4k} + \\ &+ \alpha_5 (x_{1k} * x_{2k}) + \alpha_6 (x_{1k} * x_{3k}) + \alpha_7 (x_{1k} * x_{4k}) + \alpha_8 (x_{2k} * x_{3k}) + \alpha_9 (x_{2k} * x_{4k}) + \alpha_{10} (x_{3k} * x_{4k}) \end{aligned}$$

A coefficient with a negative (positive) sign means that the respective variable contributes to a lower (higher) prediction error.

Since the most important application of labour supply models is the evaluation of tax and welfare policy reforms, we focus on the prediction performance under alternative tax regimes. More precisely, the steps above are repeated twice, with reference to the prediction of the outcomes under the current tax regime and to the prediction of the outcomes after the introduction of a flat tax.

Appendix B (Tables B.1 – B.6) reports, for the true model and for the 16 alternative models, the detailed predictions (by income decile) of participation rates, hours of work and net income, both under the current (1994) tax rule (in-sample predictions) and under the hypothetical flat tax reform (out-of-sample predictions). The results show that the introduction of a flat tax stimulates labour supply, and that the strongest labour supply response comes from females in the lower income deciles. Referring to the true model we find that the participation rates increase from 11 and 10 per cent in the

two lowest deciles to 5 per cent in the third decile. For the remaining deciles the rise in participation is rather modest. Changes in hours of work show a similar pattern as for the changes in the participation rates; i.e. the change in hours of work decreases with increasing decile. However, although labour supply of females in the richest deciles are only slightly affected by the flat tax reform these females experience a substantial increase in disposable income, which is actually larger than what can be observed for the lowest deciles.

The results of the first prediction performance regression are reported in Table 5.1. Besides reporting coefficients we also compute $100(\exp(\alpha_i) - 1)$, which measures the percentage change in the relative prediction error (i.e. z) when the variable associated to α_i changes from 0 to 1. In the notes to Table 5.1 we also provide the value of z when all the variables are set equal to 0 (which correspond to Model Ia). The estimates suggest that using a sampled alternative procedure and introducing job and peaks dummies contribute to a lower prediction error. However, the only statistically significant characteristic is *"Job dummy"*"24 alternatives"*. Overall the evidence of an important impact of alternative modes of representing the choice set as long as the replication of current values is concerned, is not strong.

In the second prediction performance exercise, the models are run after a hypothetical tax reform. A fixed proportional tax (Flat Tax) replaces the current tax system. The flat tax is determined running iteratively the "true" model until the total tax revenue is the same as under the current system. Next, the "true" outcomes (hours and net disposable income) are compared to the outcomes simulated by the 16 models and the corresponding values of the z_k are computed. When it comes to reform simulations rather than current values replication, the differences in outcomes are more marked. Table 5.2 is analogous to Table 5.1, but it refers to post-Flat-Tax outcomes. In this case we get a much clearer pattern of the effects of the different modelling strategies, in particular on the prediction of hours of work and net income. For example, when all the variables are set equal to 0 (i.e. we use Model Ia), hours of work are predicted with a relative error equal to 0.209. If we adopt sampled alternatives instead of fixed alternatives (i.e. we use Model IIIa) the relative prediction error is reduced by 83%. As follows from the detailed information provided by Tables B.4 – B.6 the less satisfactory out-of-sample prediction performance arises from discrepancies between the lower parts of the predicted and the "observed" flat tax distributions of hours of work and disposable income.

Table 5.1. Estimates of the prediction performance regression under the current tax regime

Variable	Participation probability		Hours of work		Net income	
	Coefficient α	% change in relative prediction error (z)*	Coefficient α	% change in relative prediction error (z)**	Coefficient α	% change in relative prediction error (z)***
Constant	*-1.444*		*-1.606*		*-4.153*	
Sampled alternatives	-0.291	-25.3	-0.397	-32.8	-0.435	-35.3
24 alternatives	0.638	89.3	0.400	49.2	0.440	55.3
Job dummy	-0.043	-4.23	-0.554	-42.5	-0.135	-12.63
Peaks dummy	0.159	17.23	-0.422	-34.4	-0.232	-20.71
Sampled alternatives*24 alternatives	0.541	71.8	0.589	80.2	0.369	44.63
Sampled alternatives*Job dummy	-0.890	-58.9	-0.388	-32.2	-0.156	-14.4
Sampled alternatives*Peaks dummies	0.049	5.02	-0.118	-11.1	0.094	9.9
24 alternatives*Job dummy	*-1.736*	*-82.4*	*-1.103*	*-66.8*	*-0.854*	*-57.4*
24 alternatives*Peaks dummies	0.089	9.3	0.239	27.0	0.300	35.0
Job dummy*Peaks dummies	-0.111	-10.5	0.132	14.1	-0.027	-2.67
R^2	0.877		0.879		0.823	

Notes to the Table: Coefficients in bold italics are statistically significant (< 10%).
* The relative prediction error when all the variables are zero (Model Ia) is 0.236
** The relative prediction error when all the variables are zero (Model Ia) is 0.201
*** The relative prediction error when all the variables are zero (Model Ia) is 0.016

Table 5.2. Estimates of the prediction performance regression under a flat tax reform

Variable	Participation probability		Hours of work		Net income	
	Coefficient α	% change in relative prediction error (z)	Coefficient α	% change in relative prediction error (z)	Coefficient α	% change in relative prediction error (z)
Constant	*-1.729*		*-1.566*		*-1.773*	
Sampled alternatives	-0.524	-40.8	*-0.757*	*-83.0*	*-0.238*	*-21.2*
24 alternatives	0.538	71.3	*-0.358*	*-21.2*	*-0.193*	*-17.6*
Job dummy	0.290	33.6	*-0.079*	*-17.6*	*-0.308*	*-26.5*
Peaks dummy	0.189	20.8	*-0.072*	*-26.5*	*-0.348*	*-29.4*
Sampled alternatives*24 alternatives	0.473	60.5	*0.352*	*-29.4*	*0.096*	*10.1*
Sampled alternatives*Job dummy	-0.716	-51.3	*0.120*	*10.1*	0.040	4.1
Sampled alternatives*Peaks dummies	0.055	5.7	*0.174*	*4.1*	0.032	3.3
24 alternatives*Job dummy	*-1.394*	*-75.2*	0.019	*3.3*	*0.136*	*14.6*
24 alternatives*Peaks dummies	0.122	13.0	-0.003	-14.6	0.034	3.5
Job dummy*Peaks dummies	-0.178	-16.3	*0.082*	*3.5*	*0.291*	*33.8*
R^2	0.862		0.996		0.972	

Note to the Table: Coefficients in bold italics are statistically significant (< 10%).
* The relative prediction error when all the variables are zero (Model Ia) is 0.177
** The relative prediction error when all the variables are zero (Model Ia) is 0.209
*** The relative prediction error when all the variables are zero (Model Ia) is 0.170

6. Conclusions

We have performed a series of simulation exercises aimed at exploring the performance of different versions of a labour supply model, where different approaches to represent choice sets are used. We first performed a Monte Carlo exercise where we simulate the distribution of the prediction errors of the different types of model. Since the results show that there is no notable difference among models as to the standard deviation of the prediction error distribution, we also perform a second exercise where we focus on the mean of the prediction error distribution and estimate how it is affected by different designs of the choice set representation. In this second exercise the various models are estimated using a large sample generated by a "true" model, to which they can then be compared.

The results we have obtained are likely to be application-specific rather than general, yet they produce useful suggestions. It turns out that as far as the replication of the current-tax-regime outcomes are concerned, there is little statistically significant evidence for important effects of alternative choice-set-representation procedures. Almost all the models predict well, although there are some indications favouring the sampled-alternatives procedure. However, when it comes to predicting the effect of a flat-tax reform, the indications are definitely more clear-cut. Using sampled alternatives and accounting for heterogeneity of opportunities seem to significantly reduce the prediction errors.

The simulation experiments illustrated in this paper suggest that indeed the issues related to the representation of the choice set in the discrete choice framework are worthwhile a more attentive design than it is commonly done in the literature on labour supply. This seems especially relevant in view of using the models for the prediction of policy effects. The prediction performance of current values does not significantly discriminate between different models, but the prediction performance of a post-reform does. These results convey the important message that the ability of a model to replicate observed outcomes is not very informative. Ultimately, the models and the procedures used to develop them should be judged on their ability to do the job they are built for, i.e. predicting the outcomes of policy changes.

Appendix A

Here we report the parameter estimates of the "true" model and of the 16 alternative models.

Table A.1. The "true" model

	Parameter	Estimate	Standard Dev.
Utility function			
Consumption			
	α_1	0.39	0.11
	α_2	4.42	0.44
Leisure			
	α_3	-4.57	0.53
	α_4	168.88	27.47
Log age	α_5	-94.29	15.32
Log age squared	α_6	13.35	2.16
Number of children below 3 years old	α_7	0.44	0.23
Number of children 3-6 years old	α_8	1.23	0.24
Number of children 7-14 years old	α_9	1.05	0.19
Opportunity density			
Job dummy	θ_0	-0.60	0.10
Part-time dummy	π_1	0.46	0.10
Full-time dummy	π_2	1.57	0.07

Table A.2. Estimates of the 16 alternative models

	True model	Model Ia	Model IIa	Model Ib	Model IIb	Model Ic	Model IIc	Model Id	Model IId	Model IIIa	Model IVa	Model IIIb	Model IVb	Model IIIc	Model IVc	Model IIId	Model IVd
α_1	**0.39**	0.35	0.54	0.43	0.46	0.43	0.50	0.43	0.44	0.54	0.55	0.53	0.54	0.55	0.55	0.52	0.53
α_2	**4.42**	2.46	3.70	3.97	4.55	4.05	4.64	4.17	4.38	3.96	3.93	4.72	4.64	4.56	4.51	4.70	4.62
α_3	**-4.57**	-7.53	-3.18	-7.31	-6.72	-2.07	-0.14	-3.99	-4.15	-5.15	-5.27	-5.94	-6.10	-2.40	-2.49	-3.52	-3.60
α_4	**168.88**	54.20	184.85	64.76	92.39	232.99	351.30	156.91	171.12	125.90	121.50	112.19	106.31	234.88	231.26	195.26	190.72
α_5	**-94.29**	-30.46	-102.83	-36.27	-51.64	-128.78	-193.30	-87.38	-95.45	-70.17	-67.75	-62.54	-59.28	-129.94	-128.03	-108.43	-105.95
α_6	**13.35**	4.32	14.62	5.15	7.33	18.27	27.48	12.40	13.54	9.96	9.62	8.88	8.42	18.46	18.19	15.39	15.04
α_7	**0.44**	0.13	0.51	0.13	0.19	0.61	0.95	0.38	0.40	0.33	0.30	0.25	0.23	0.66	0.58	0.50	0.44
α_8	**1.23**	0.48	1.68	0.53	0.76	1.86	2.99	1.25	1.40	1.07	1.05	0.91	0.87	1.94	1.95	1.56	1.57
α_9	**1.05**	0.40	1.37	0.44	0.62	1.53	2.47	1.04	1.14	0.88	0.88	0.75	0.73	1.61	1.65	1.29	1.33
θ_0	**-0.60**	-	-	-1.08	-2.33	-	-	-0.78	-2.10	-	-	-0.88	-0.86	-	-	-0.63	-0.60
π_1	**0.46**	-	-	-	-	-0.23	0.14	0.15	0.28	-	-	-	-	0.44	0.44	0.53	0.52
π_2	**1.57**	-	-	-	-	0.99	1.53	0.78	1.19	-	-	-	-	1.66	1.63	1.56	1.54

Appendix B

Note that "income decile" in Tables B.1 – B.6 refers to the distribution of disposable income (income after tax) as predicted by the different models under the 1994 tax system.

Table B.1 Examples of predictions of participation rates under the 1994 tax system.

Income Decile	True model	Model Ia	Model Id	Model IIa	Model IId
1	**58**	55	87	87	*55*
2	**65**	73	93	93	*67*
3	**79**	81	95	95	*79*
4	**86**	87	97	97	*85*
5	**91**	92	96	96	*90*
6	**93**	94	98	98	*93*
7	**93**	95	99	99	*94*
8	**94**	94	98	98	*93*
9	**94**	95	99	99	*96*
10	**88**	89	97	97	*87*
Mean	**84**	86	96	87	*84*

Table B.2 Examples of predictions of hours of work under the 1994 tax system.

Income Decile	True model	Model Ia	Model Id	Model IIa	Model IId
1	**568**	627	*514*	733	*568*
2	**715**	818	*730*	837	*708*
3	**937**	1000	*890*	989	*941*
4	**1157**	1179	*1130*	1125	*1153*
5	**1389**	1375	*1397*	1276	*1352*
6	**1527**	1494	*1541*	1429	*1528*
7	**1606**	1638	*1650*	1598	*1631*
8	**1695**	1701	*1735*	1667	*1672*
9	**1757**	1812	*1838*	1746	*1771*
10	**1523**	1631	*1566*	1676	*1567*
Mean	**1287**	1327	*1299*	1308	*1289*

Table B.3 Examples of prediction of disposable income (in NOK) under the 1994 tax system.

Income Decile	True model	Model Ia	Model Id	Model IIa	Model IId
1	**168915**	170648	*169098*	171945	*168690*
2	**216080**	217801	*215357*	219415	*216333*
3	**244914**	245504	*243740*	245176	*243672*
4	**268880**	268308	*267340*	267880	*267659*
5	**290441**	290083	*290556*	288798	*289893*
6	**312088**	312113	*313719*	310410	*312446*
7	**336247**	335829	*337305*	334374	*336148*
8	**363833**	364607	*365453*	362513	*363739*
9	**403513**	405063	*405654*	403401	*404046*
10	**600841**	605283	*602163*	608705	*604516*
Mean	**320575**	321524	*321038*	321262	*320714*

Table B.4 Examples of predictions of participation rates under a flat tax reform

Income Decile	True model	Model Ia	Model Id	Model IIa	Model IId	Model IIIa	Model IIId	Model IVa	Model IVd
1	**69**	62	55	89	63	76	65	76	65
2	**75**	77	68	95	74	83	74	84	75
3	**84**	83	76	96	83	90	83	90	83
4	**89**	89	83	97	87	92	88	93	89
5	**93**	93	89	97	91	94	91	94	91
6	**94**	94	92	98	93	95	94	95	94
7	**94**	95	92	99	94	95	93	96	93
8	**95**	94	92	98	93	97	96	97	96
9	**95**	96	94	99	96	98	96	98	96
10	**88**	89	84	97	88	92	88	92	88
Mean	**88**	87	82	96	86	91	87	91	87

Table B.5 Examples of predictions of hours of work under a flat tax reform

Income Decile	True model	Model Ia	Model Id	Model IIa	Model IId	Model IIIa	Model IIId	Model IVa	Model IVd
1	**987**	835	826	946	890	890	883	921	880
2	**1022**	943	966	1041	943	943	993	985	992
3	**1160**	1100	1117	1145	1134	1134	1131	1151	1133
4	**1315**	1260	1279	1271	1291	1291	1330	1307	1338
5	**1491**	1432	1488	1392	1459	1459	1493	1460	1485
6	**1609**	1542	1626	1543	1609	1609	1650	1579	1646
7	**1659**	1677	1717	1685	1670	1670	1691	1675	1695
8	**1742**	1735	1786	1727	1720	1720	1775	1771	1774
9	**1794**	1843	1898	1811	1821	1821	1811	1807	1814
10	**1549**	1647	1619	1721	1606	1606	1587	1617	1586
Mean	**1487**	1401	1432	1428	1414	1414	1434	1427	1434

Table B.6 Examples of predictions of net income under a flat tax reform

Income Decile	True model	Model Ia	Model Id	Model IIa	Model IId	Model IIIa	Model IIId	Model IVa	Model IVd
1	**194076**	171081	177612	173092	177934	175360	178959	175829	178558
2	**234263**	214268	220564	222704	220524	221008	223384	220745	222943
3	**259189**	242704	250457	247374	248492	248332	249373	247584	249304
4	**279624**	266384	272361	271441	271579	272276	275414	273516	275739
5	**301124**	289038	294062	293453	294681	293241	296123	293368	295567
6	**323777**	314124	320755	319278	319492	318317	321883	318698	321400
7	**350809**	342509	349310	346358	344397	346147	348328	346124	348868
8	**383958**	375740	379893	378941	377972	377469	379296	378295	378984
9	**431297**	426513	431747	430622	428668	430380	430587	429954	431015
10	**651815**	649764	651885	657771	652667	651514	650805	652383	650766
Mean	**340993**	329213	334865	334103	333641	333404	335415	333650	335314

Aknowledgements

Notes

[1] See also Creedy and Kalb (2004) for a survey of alternative approaches to discrete labour supply.

[2] A random variable ε has a (standard) Type I extreme value distribution if Prob($\varepsilon \leq k$) $= \exp(-\exp(-k))$.

[3] There is still another approach, the so called Dogit model (Gaudry and Dagenais, 1979), to represent a non-uniform availability of alternatives. It is a generalization of the logit model, where the decision-maker may – with a given probability – be "captive" to one of the alternatives or otherwise choose freely from the whole choice set. The Dogit model has been recently used by Harris and Duncan (2002) in a labour supply application. We do not consider the Dogit model in the simulation exercise presented here.

[4] See e.g. Aaberge, Colombino and Strøm, (1999, 2000, 2004) and Aaberge, Colombino, Strøm and Wennemo (2000).

[5] Since the density values must add up to 1, we can also compute γ according to:

$$\gamma\big((910-52)+(1066-52)\big)\exp(\pi_1)+(1898-1066)+(2106-1898)\exp(\pi_2)+(3640-2106)\big)=1$$

[6] For the sake of simplicity we presented the model as if the wage rates were given and exogenous. However, as a matter of fact the wages are treated as endogenous and the wage functions are simultaneously estimated by maximum likelihood together with the utility function and the job opportunity density.

[7] In the estimation of these model, the actual wage rates are used for the working individuals, while to non-working individual we impute wages computed on the basis of a wage equation estimated with a two-step-Heckman procedure. The estimates of the wage equation are available upon request from the authors.

[8] The results for the other models (available upon request from the authors) do not add significant evidence beyond what is revealed by the four models we focus upon.

[9] Since z is always positive, a linear specification would not be appropriate. We therefore use ln(z).

References

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.

Aaberge, R., U. Colombino and S. Strøm (1999) Labor Supply in Italy An Empirical Analysis of Joint Household Decisions, with Taxes and Quantity Constraints, *Journal of Applied Econometrics*, 14, 403-422.

Aaberge, R., U. Colombino and S. Strøm (2000) Labour Supply Responses and Welfare Effects from Replacing Current Tax Rules by a Flat Tax : Empirical Evidence from Italy, Norway and Sweden, *Journal of Population Economics,* 13, 595-621.

Aaberge, R., Colombino U., Strøm S. and T. Wennemo (2000) Joint Labour Supply of Married Couples: Efficiency and Distribuitional Effects of Tax Reforms, in: Sutherland, Mitton, Sutherland and Weeks (eds.) *Microsimulation Modelling for Policy Analysis: Challenges and Innovations*, Cambridge University, Cambridge, U.K.

Aaberge, R., U. Colombino and S. Strøm (2004) Do More Equal Slices Shrink the Cake? An Empirical Investigation of Tax-Transfer Reform Proposals in Italy, *Journal of Population Economics*, 17, 767-785.

Ben-Akiva, M. and T. Watanatada (1981) Application of a Continuous Spatial Choice Logit Model, in Manski, C. F. and McFadden D. (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.

Ben-Akiva, M., and S. R. Lerman (1985) *Discrete choice analysis*, (MIT Press, Cambridge).

Blomquist, S. (1988) Non-linear Taxes and Labor Supply, *European Economic Review*, 32, 1213-1226.

Blundell, R. and T. MaCurdy (2000) Labor Supply: A Review of Alternative Approaches, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Elsevier North-Holland.Blundell, R., A. Duncan, A., J. McCrae and C. Meghir (2000) The Labour Market Impact of the Working Families' Tax Credit, *Fiscal Studies*, 21, 75-100.

Colombino U. (1998) Evaluating the effects of new telephone tariffs on residential users' demand and welfare. A model for Italy, *Information Economics and Policy*, 10, 283-303.

Creedy, J. and G. Kalb (2005) Discrete Hours Labour Supply Modelling: Specification, Estimation and Simulation, *Journal of Economic Surveys*, 19, 97-734

Dagsvik, J.K. (1994) Discrete and Continuous Choice, Max-stable Processes and Independence from Irrelevant Attributes, *Econometrica*, 62, 1179-1205.

Duncan, A. and M. Weeks (1997) Behavioural Tax Microsimulation with Finite Hours Choices, *European Economic Review*, 41, 619-626.

Gaudry, M. and M. Dagenais (1979) The Dogit Model, *Transportation Research*, 13B, 105 – 112.

Harris, M.N. and A. Duncan (2002) Intransigencies in the Labour Supply Choice, Melbourne Institute Working Paper No 17/02.

Hausman, J.A. (1979) The Econometrics of Labour Supply on Convex Budget Sets, *Economic Letters*, 3, 171-174.

Heckman, J. (1974) Effects of Child-Care Programs on Women's Work Effort, *Journal of Political Economy*, 82, 136-163.

Kornstad, T. and T.O. Thoresen (2004) Means-testing the Child Benefit, *Review of Income and Wealth*, 50, 29-49.

MaCurdy, T., D. Green and H. Paarsch (1990) Assessing Empirical Approaches for Analyzing Taxes and Labor Supply, *Journal of Human Resources,* 25, 415–449.

McFadden, D. (1974) Conditional Logit Analysis of Qualitative Choice Behavior, in P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York.

McFadden, D. (1978) Modelling the Choice of Residential Location in A. Karlquist, L. Lundquist, F. Snickard and J.J. Weilbull (eds.): *Spatial Interaction Theory and Planning Models*, Amsterdam, North-Holland.

McFadden, D. (1981) Structural Discrete Probability Models Derived from Theories of Choice, in Manski, C. F. and McFadden D. (eds.) *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.

Moffitt, R. (1986) The Econometrics of Piecewise-Linear Budget Constraints: A Survey and Exposition of the Maximum Likelihood Method, *Journal of Business & Economic Statistics*, 4, 317-328.

Train, K.E., D. L. McFadden and M. Ben-Akiva (1987) The demand for local telephone service, *Rand Journal of Economics*, 18, 109-123.

van Soest, A. (1995) Structural Models of Family Labor Supply: A Discrete Choice Approach, *Journal of Human Resources*, 30, 63-88.

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.