The estimation of utility consistent labor supply models by means of simulated scores

Bloemen, H.G.; Kapteyn, A.J.

Publication date:
1993

entER
for
omic Research

Discussion paper

[redacted]

TILBURG UNIVERSITY

Center
for
Economic Research

No. 9367

THE ESTIMATION OF UTILITY CONSISTENT LABOR SUPPLY MODELS BY MEANS OF SIMULATED SCORES

by Hans G. Bloemen and
Arie Kapteyn

October 1993

ISSN 0924-7815

THE ESTIMATION OF UTILITY CONSISTENT LABOR SUPPLY MODELS BY MEANS OF SIMULATED SCORES

by

Hans G. Bloemen and Arie Kapteyn [1]

CentER for Economic Research
Tilburg University
P.O. BOX 90153
5000 LE Tilburg
The Netherlands

First version October 1992
This version September 1993

Abstract

We specify a utility consistent static labor supply model with flexible preferences, a non-linear and possibly non-convex budget set, and a wage equation. Three stochastic error terms are introduced to represent respectively optimization and reporting errors, stochastic preferences, and heterogeneity in wages. Coherency conditions on parameters and supports of error distributions are imposed for all observations. The complexity of the model makes it impossible to write down the probability of participation. Hence simulation techniques have to be used in estimation. The properties of the estimation method adopted are first investigated by means of Monte Carlo. After that the model is estimated for Dutch data.

We compare our approach with various simpler alternatives proposed in the literature. It turns out that both in the Monte Carlo and for the real data the various estimation methods yield very different results. Since, moreover, our estimation method yields good results for the Monte Carlo data, we suggest that the simplifications adopted in the literature may have generated considerable biases.

[1] The authors thank the Organisatie voor Strategisch Arbeidsmarktonderzoek (OSA) for kindly providing the data. Thanks are due to Arthur van Soest for his help and comments.

1 Introduction

By now there is an enormous literature on the estimation of static models of individual labor supply. Typically, a model will consist of two equations, a wage equation and a labor supply equation. Especially since the work of Hausman (1979, 1980, 1985) the labor supply equation is usually utility consistent[2] and often the underlying budget set is piecewise linear and possibly non-convex. See e.g. Blomquist (1983), Moffit (1986) and the papers in the special issue of the *Journal of Human Resources*, Summer 1990.

Despite the vast quantity of papers written on the topic, there are still various unsatisfactory elements in the models estimated so far. These pertain to both the specification and the estimation of the models. As to specification, one usually adheres to simple forms of the labor supply function, whereas the stochastic specification is often more governed by considerations of convenience than of plausibility. Estimation of simple models is not much of a problem (e.g. of a type II Tobit model), but in somewhat more complicated models often short cuts are being taken that strictly speaking impair consistency of estimators. In the next section these issues will be discussed in more detail. There are good reasons for all of this. As we will illustrate in the next section, essentially the canonical Hausman model with a flexible specification of preferences, a non-convex budget set and a proper stochastic specification could not be estimated with methods available until recently. Possibly the most glaring difficulty is that except in very simple models it is impossible to write down the probability of participation. Since this probability plays a role in any estimation method one would like to apply, ranging from ML to MM, all estimation methods applied in practice can be seen as approximations with varying degrees of accuracy. [3]

Rather closely related to the previous issues is the issue of coherency. [4] It turns out that in models with kinked budget constraints coherency requires quasiconcvity of the direct utility function at all kink points. Since these kink points will be different for different individuals in a sample, coherency requires quasiconcavity at many combinations of hours and wages. See MaCurdy, Green, Paarsch (1990) or Van Soest, Kooreman, and Kapteyn (1993). This in turn means that parameters and error distributions have to be restricted in order to make sure that the model is utility consistent (i.e., the direct utility function is strictly quasi-concave) at relevant kink points for each observation. [5] Except for simple models, the imposition of utility consistency is non-trivial. Failure to do so, however, may lead to inconsistent estimators.

[2] By "utility consistent" we mean throughout this paper that observed or predicted labor supply can be rationalized as the result of the maximization of a well- behaved utility function; we call a utility function well- behaved if it is strictly quasi-concave and increasing in consumption

[3] In this paper we do not pay attention to more extensive models where the hours decision and participation decision are modelled separately, as in e.g. Blundell and Meghir (1987) or Blundell, Ham, and Meghir (1987). Although at first sight this may seem to circumvent the problem mentioned, a fully consistent treatment will still require the computation of the probability that desired hours are zero, and that is precisely the problem we are dealing with.

[4] See e.g Gourieroux, Laffont, and Montfort (1980). A model is coherent if endogenous variables are uniquely determined by the exogenous variables and the errors.

[5] If one allows for measurement error, the construction of a likelihood requires that one integrates over all possible values of the "true" number of hours, which includes all kink points.

In this paper we specify a utility consistent static model of individual labor supply, with a flexible specification of preferences. The model is of the conventional two-equation form, a wage equation and a labor supply equation. Three random errors are specified, one additive random error in the wage equation representing unobserved heterogeneity, one additive error in the labor supply equation representing optimization and reporting errors and a non-additive error in the labor supply equation representing random preferences. We impose utility consistency at all data points. The estimation method is a variant of a method of simulated scores (MSS), developed in an earlier paper (Bloemen and Kapteyn, 1993). Thereby we avoid the impossible task of writing down the probability of participation; instead we draw from the errors in the model and simply determine whether utility is maximal while working or not working. Since the model and estimation method are rather intricate, we first look at the properties of the estimation method by means of Monte Carlo. Next the model is estimated for Dutch data on married females. In recent years substantial advances have been made in the development of computationally efficient simulators. See e.g. the survey by Hajivassiliou (1991). All these approaches exploit to some extent specific properties of the model at hand, like linearity, normality, or smoothness. In the present context none of these properties applies. Hence the estimation method will be rather brute force, using frequency simulators and numerical approximations of derivatives whenever required.

The order of presentation is as follows: In the next section we set out the basic model and discuss various approaches in the literature using the model as an illustration. In section 3 the estimator is presented. In section 4 we give a detailed specification of the model and the restrictions that have to be imposed to render the model coherent. In section 5 we present details of the simulation methods needed to operationalize the estimator. In section 6 we compare our estimator with a number of alternatives used in the literature. We first do this for artificial data generated by means of Monte Carlo and next for real data. Section 7 concludes.

2 The economic model

Consider the following utility function, which is a special case of the utility function proposed by Hausman and Ruud (1984):

$$U(h,c) = m^* \exp\left\{\frac{\beta}{\gamma}(h-\delta-\beta m^*)\right\} \tag{2.1}$$

where

$$m^* = \frac{\gamma}{\beta^2}\left[1-\left\{1+\frac{\beta^2}{\gamma}\left[\frac{(h-\delta)^2}{\gamma}-2(c+\theta)\right]\right\}^{\frac{1}{2}}\right] \tag{2.2}$$

The variables h and c are hours worked and consumption respectively; $\beta, \gamma, \delta, \theta$ are parameters. Maximization of this utility function subject to a linear budget constraint of the form $c \leq wh + \mu$ yields the following labor supply function:

$$h(w,\mu) = \delta + \mu^* \beta + w\gamma \tag{2.3}$$

with

$$\mu^* = \theta + \mu + \delta w + \frac{1}{2}\gamma w^2 \tag{2.4}$$

The cost function dual to the utility function is:

$$c(u, w, \mu) = u.\exp(\beta w) - \{\theta + \delta w + \frac{1}{2}\gamma w^2\} \tag{2.5}$$

We will assume throughout that $\gamma > 0$. It is easy to see that concavity of the cost function requires

$$\mu^* < \gamma/\beta^2, \tag{2.6}$$

An equivalent condition in (c, h) space is that the indifference curves are convex. This requires:

$$m^* < \gamma/\beta^2 \tag{2.7}$$

Notice that under this condition utility is increasing in consumption.

Using (2.2), (2.7) can be written as follows:

$$\left\{1 + \frac{\beta^2}{\gamma}\left[\frac{(h-\delta)^2}{\gamma} - 2(c+\theta)\right]\right\}^{\frac{1}{2}} > 0 \tag{2.8}$$

Clearly this holds true whenever the utility function is defined. The condition for the existence of the utility function can be written as:

$$\frac{(h-\delta)^2}{\gamma} - 2(c+\theta) > -\frac{\gamma}{\beta^2} \tag{2.9}$$

or,

$$c < -\theta + \frac{\gamma}{2\beta^2} + \frac{(h-\delta)^2}{2\gamma} := f(h) \tag{2.10}$$

with $f(h)$ implicitly defined. The function $f(h)$ is a parabola with a minimum for $h = \delta$. The value of the minimum is $-\theta + \gamma/2\beta^2$. Figure 1 sketches the domain of $U(h, c)$:

Let us now turn to the description of behavior under a nonlinear (actually piecewise linear) and non-convex budget set. Figure 2 represents the familiar example of a utility maximum attained at the point where an indifference curve is tangent to the budget constraint.

As the budget set is not convex, but can be seen as the union of convex sets, an algorithm for finding the utility maximum is to first find points of tangency or corner solutions (kink points) for each convex set and then picking the point which yields the maximum maximorum.

To complete the model we need an equation explaining the before tax wage of an individual and the specification of the stochastic structure. Furthermore, we introduce a subscript n to index the observations, $n = 1, \ldots, N$. The wage equation is specified as follows:

$$\log w_n = \sum_{j=1}^{J} \eta_j x_{nj} + u_n \tag{2.11}$$

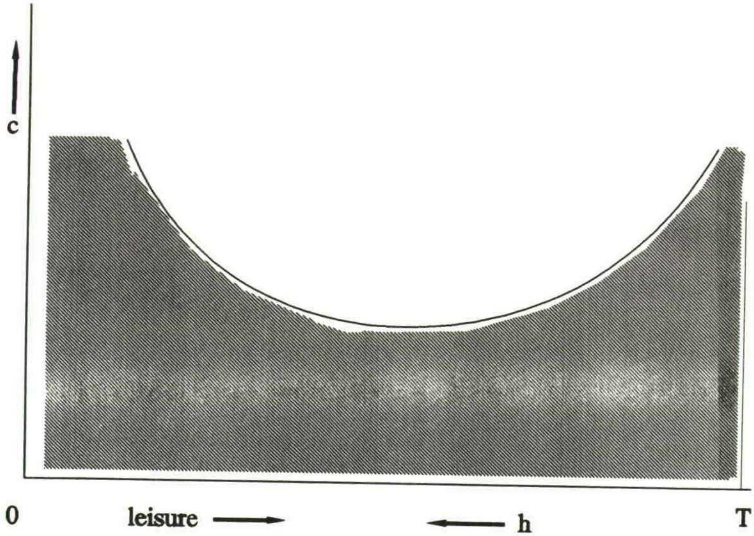

Figure 1: The domain of U(h,c)

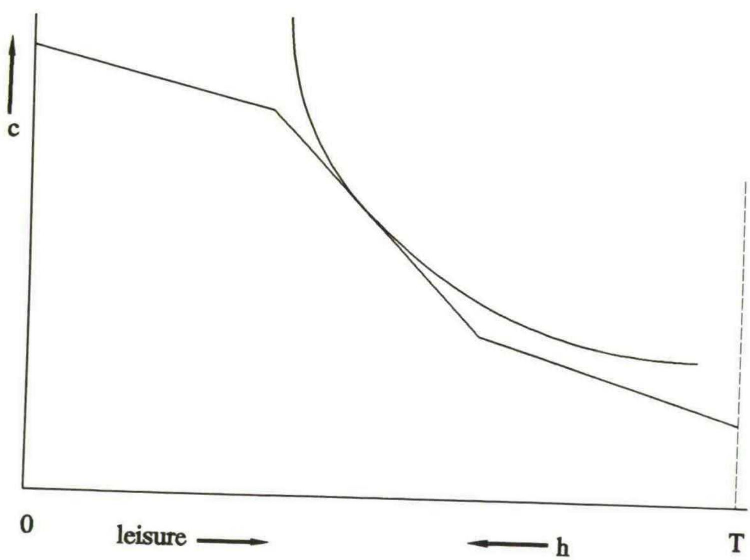

Figure 2: Utility maximization over a nonconvex budget set

where u_n is an i.i.d. error term representing unobserved heterogeneity, x_{nj} are observable characteristics and η_j are parameters. As to the labor supply equation, we introduce preference variation by allowing θ to vary across agents as follows:

$$\theta_n = \theta_0 + x_n'\omega + v_n \tag{2.12}$$

where x_n is a vector of individual observable characteristics and v_n an unobservable error term; ω is a parameter vector.[6] For later purposes it is useful to define also

$$\mu_n^* = \theta_n + \mu_n + w_n\delta + \frac{1}{2}w_n^2\gamma \tag{2.13}$$

For given values of w_n and v_n indivdual n's optimal number of hours, say $\overline{h}_n$ is determined as in Figure 2. In general this is a complicated function of the wage, nonlabor income, individual characteristics and the random preference term. We write this as

$$\overline{h}_n = h_n(w_n, \mu_n; \alpha, x_n'\omega + v_n), \tag{2.14}$$

where $\alpha = (\beta, \gamma, \delta, \theta_0)'$. Notice for instance that the function $h_n(.)$ need not be continuous. We allow for the possibility of optimization or measurement errors by adding an error term ϵ_n:

$$h_n^* = \overline{h}_n + \epsilon_n \tag{2.15}$$

Let h_n be observed labor supply, then we assume:

$$h_n = h_n^* \quad \text{if } h_n^* > 0 \quad \underline{\text{and}}\ h_n(w_n, \mu_n; \alpha, x_n'\omega + v_n) > 0 \tag{2.16}$$

$$h_n = 0 \quad \text{if } h_n^* \leq 0 \quad \underline{\text{or}}\ h_n(w_n, \mu_n; \alpha, x_n'\omega + v_n) \leq 0 \tag{2.17}$$

This formulation brings out the distinction between the random preferences v_n and the optimization or measurement errors ϵ_n. This type of stochastic specification is in line with work of Hausman (1981).

With this model at hand, we can discuss various problems in estimation and characterize different approaches in the literature.

- *It is difficult to derive the density of wages and hours for working individuals if the budget constraint is non- convex.* The joint density of hours and (before tax) wages can be written as the product of the conditional density of hours given wages and the marginal density of wages. The latter is not difficult to write down, but the former may be. As indicated above, if the budget constraint is non-convex it can be written as the union of convex sets and the obvious thing to do is to first find the utility maximum in each convex set and next compute the *maximum maximorum.* With random preferences this means that we have to find the density of hours for each convex subset and the probability that the *maximum maximorum* is in any of the convex sets. Finding the density of hours for each convex set is tedious but feasible (see section 5), but the probability that the utility maximum occurs in any given convex subset is almost impossible to write down as one can easily imagine by inspecting the formula for the direct utility function. This difficulty will arise in all but the simplest utility specifications.

[6] There is no a priori reason to let preference variation enter through θ only, any of the other parameters of the utility function may be made dependent on observable and unobservable characteristics. For simplicity of the exposition we stick to the present somewhat arbitrary choice

- *It is very hard to write down the probability of participation.* To write down the probability of participation, one has to characterize the values of ϵ_n, u_n, v_n for which individual n will be observed working. For non-working individuals the wage they could earn while working is typically not known. The random variables u_n and v_n cause the budget constraints and the indifference curves to move around in a complicated way. Even for a given budget constraint (i.e. for someone who does participate) it is difficult to find the values of v_n for which the utility of working will exceed that of not working for the same reason as given above. Since the budget constraint is the result of the interplay of the gross wage with possibly quite complicated institutions, the resulting distribution of the budget constraint will in general be intractable. Combined with the difficulty of writing down the probability of working for a given budget constraint, this makes it impossible to write down the probability of participation as an analytic function of exogenous variables and parameters.

- *Incoherency.* The problem of finding a utility maximum is generally well-defined if indifference curves are convex. However, if a flexible specification is adopted for the utility function, it will generally not be globally quasi-concave and hence there will be combinations of the parameters and values of exogenous variables and errors for which indifference curves are not convex or are not defined, cf. Fig. 1. As shown by Van Soest, Kooreman, Kapteyn (1993), this means that the model is no longer coherent. This in turn implies that estimation methods are not well-defined. To have well-defined estimation, coherency has to be imposed. For instance, these authors give an example where data are generated by a coherent model, but no coherency is imposed in estimation. It is shown that in that case the "likelihood" does not attain its maximum at the true parameter point, but rather at a point which violates coherency.

- *Time consuming numerical integration.* Even if we are able to write down in principle the probability of certain events or the density of wages and hours, they are bound to be complicated expressions involving multi-dimensional integrals. Since the model involves various non-linear transformations it is very unlikely that analytical solution of the integrals is possible. Numerical integration tends to be extremely time consuming.

To solve or evade these problems, various routes can be taken.

- *Choose simple functional forms.* As said above, to write down the joint density of (before tax) wages and hours for working individuals one needs the density of hours conditional on the wage times the marginal density of the wage. The latter is straightforward. The former can be simplified considerably by choosing a simple specification, like e.g. the Hausman linear labor supply function. This function arises from model (2.1)- (2.4) by letting γ approach zero. The utility function then reduces to
$$U(h,c) = \frac{\delta - h}{\beta} \exp\left\{-1 + \frac{\beta(c+\theta)}{h-\delta}\right\} \tag{2.18}$$

- *Ignore random preferences.* This can take two forms. One can ignore random preferences altogether, so that the only source of random variation of hours given wages is optimization or reporting errors (see e.g. Kooreman and Kapteyn (1986)). Alternatively, sometimes a random error is appended to the (non-stochastic) utility difference between working and not working, as in Kapteyn, Kooreman and Van Soest(1990). This latter term may also have the interpretation of random preferences. It is of course a bit hard to see why preferences would only be random in utility comparisons and not in the hours choice. [7]

- *Ignore unobserved heterogeneity in wages for non- participants.* In this approach the wage equation is estimated for working individuals and next used to predict before tax wages for non-participants. The implied budget constraint is assumed to be the true budget constraint, with neglect of unobserved heterogeneity. Notice that for the estimation of the wage equation for working individuals some correction for selectivity bias is required, which in turn requires the probability of participation. Strictly speaking, one should use the full model to estimate this probability. Since, as indicated, this is either very difficult or impossible, some approximation of this probability is used at this stage.

- *Use working individuals only, with correction for selectivity.* By only using working individuals one does not really avoid the necessity of computing the probability of participation, but one can approximate this probability as in the previous approach. Often, if only working individuals are used in the analysis, the budget constraint is linearized in the observed point. Of course the marginal wage used to linearize the budget constraint is endogenous, but this may be solved by the use of instrumental variables. In this approach, typically no steps are taken to guarantee coherency of the model in all data points. For this reason. it is not quite clear whether the estimation method is consistent or not.

3 Estimation

This section largely follows Bloemen and Kapteyn (1993). For ease of notation we introduce the dummy variable d_n with

$$d_n = 1 \text{ if } h_n = 0 \tag{3.1}$$

$$d_n = 0 \text{ if } h_n > 0 \tag{3.2}$$

Furthermore we write $P_n(\vartheta)$, where ϑ contains the parameters of α, η and the parameters of the distribution functions of u_n, v_n, and ϵ_n, for the probability that h_n equals zero. The joint density of wages and hours for a participating agent n is denoted by $g^*(h_n^*, w_n | x_n, \mu_n, \vartheta)$, where

$$-\infty < h_n^* < \infty$$
$$0 < w_n < \infty$$

[7] An alternative interpretation may be that the random term in the utility comparison represents optimization errors, which may be more natural in certain contexts.

From this we can derive the mixed discrete-continuous probability density function of h_n and w_n, $g(h_n, w_n|x_n, \mu_n, \vartheta)$.

$$g(h_n, w_n|x_n, \mu_n, \vartheta) = \begin{cases} P_n(\vartheta) & \text{if } h_n = 0 \\ g^*(h_n, w_n|x_n, \mu_n, \vartheta) & \text{if } h_n > 0, 0 < w_n < \infty \end{cases}$$

For ease of notation we will often denote the probability $P_n(\vartheta)$ by P_n. We shall denote the probability of working $1 - P_n(\vartheta)$ by $\bar{P}_n(\vartheta)$ or simply by $\bar{P}_n$. We assume that our sample is ordered in such a way that the observations 1 to N_1 refer to non-working individuals and the observations $N_1 + 1$ to N are working individuals.

Generally, the log-likelihood function of the model is

$$L(\vartheta|x_n, \mu_n, w_n, h_n, n = 1, ..., N) = \sum_{n=1}^{N_1} \ln P_n(\vartheta) + \sum_{n=N_1+1}^{N} \ln g^*(h_n, w_n|x_n, \mu_n, \vartheta) \tag{3.3}$$

It will be assumed that the likelihood is differentiable almost everywhere with respect to the elements of ϑ[8]. Thus in principle we can differentiate the log-likelihood with respect to ϑ to derive the first order conditions for a maximum.

$$\frac{\partial L(\vartheta)}{\partial \vartheta} = \sum_{n=1}^{N_1} \frac{\partial \ln P_n(\vartheta)}{\partial \vartheta} + \sum_{n=N_1+1}^{N} \frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \vartheta)}{\partial \vartheta} \tag{3.4}$$

$$\frac{\partial L(\hat{\vartheta}_{ML})}{\partial \vartheta} = 0 \tag{3.5}$$

where $\hat{\vartheta}_{ML}$ is the maximum likelihood estimator of ϑ.

Alternatively, we can rewrite the derivative of the log-likelihood function as

$$\frac{\partial L(\vartheta)}{\partial \vartheta} = \sum_{n=1}^{N} \left[d_n \frac{\partial \ln P_n(\vartheta)}{\partial \vartheta} + (1 - d_n) \frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \vartheta)}{\partial \vartheta} \right] \tag{3.6}$$

where d_n is the dummy variable introduced above. Let ϑ_0 be the true parameter value. It is well known that if the support of h_n and w_n does not depend on ϑ, the score vector has expectation zero:

$$E\left(\frac{\partial L(\vartheta_0)}{\partial \vartheta} \right) = 0 \tag{3.7}$$

It is this fact which implies consistency of the ML estimator. In the present context the evaluation of the score vector is impossible for the reasons set out in the previous section. We will replace the score by an unbiased simulator, which can then still be used for consistent estimation of the parameters in ϑ.

We rewrite the first order derivative of the log-likelihood function in the following way.

$$\frac{\partial L(\vartheta)}{\partial \vartheta} = \sum_{n=1}^{N} \left\{ Z_n(d_n - P_n) + (1 - d_n) \left[\frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \vartheta)}{\partial \vartheta} - \frac{\partial \ln \bar{P}_n}{\partial \vartheta} \right] \right\} \tag{3.8}$$

[8]For consistency some additional regularity conditions are required: $\frac{\partial^2 L}{\partial \vartheta \partial \vartheta'}$ exists in a neighbourhood of ϑ_0 and is non-singular and negative definite in a neighbourhood of ϑ_0.

where

$$Z_n := \frac{\frac{\partial P_n}{\partial \vartheta}}{P_n(1-P_n)} \tag{3.9}$$

The first component of this expression equals the score of the log-likelihood of the binary response model. If we replace the vector Z_n by an arbitary vector of instruments $\bar{Z}_n$, independent of ϑ, the expectation of the resulting expression, conditional on $\bar{Z}_n$, equals zero at the true parameter value ϑ_0.

In Section 5 we will describe how we simulate $P_n(\vartheta)$ and its derivative unbiasedly. As to the second term in (3.8) the simulation of the part involving the density g^* will be described in Section 5 as well. And finally, we use the fact that

$$E\left[(1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \vartheta}\right] = \frac{\partial \bar{P}_n}{\partial \vartheta} \tag{3.10}$$

so that we may replace $(1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \vartheta}$ by $\frac{\partial \bar{P}_n}{\partial \vartheta}$ without affecting the unbiasedness property of the score. As a result, the original score vector is replaced by

$$\frac{\partial \bar{L}}{\partial \vartheta} = \sum_{n=1}^{N}\left[\bar{Z}_n(d_n - P_n) + (1-d_n)\frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \vartheta)}{\partial \vartheta} - \frac{\partial \bar{P}_n}{\partial \vartheta}\right] \tag{3.11}$$

Inserting unbiased simulators k_{nR} and $\bar{m}_{nR}$ based on R replications for the response probabilities and their derivatives respectively in this expression gives the *simulated score*:

$$K_R(\vartheta) = \sum_{n=1}^{N}\left[\bar{Z}_n(d_n - k_n(\vartheta, v_R^*)) + (1-d_n)\frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \vartheta)}{\partial \vartheta} - \bar{m}_n(\vartheta, v_R^*)\right] \tag{3.12}$$

The advantage of writing the score vector this way is that the simulators for the response probabilities and their derivatives enter the expression linearly. As a result simulation errors are averaged out over individuals. Moreover, if a frequency simulator is used to simulate the response probabilities discontinuities are averaged over individuals, thereby eliminating the reason for the poor performance of the frequency simulator in the context of simulated maximum likelihood estimation. See, e.g., Lerman and Manski (1981) and Börsch-Supan and Hajivasiliou (1993).

The estimation procedure now becomes: Choose instrument vectors $\bar{Z}_n$ and obtain the estimator by solving the moment conditions:

$$K_R(\vartheta) = 0 \tag{3.13}$$

To ascertain the efficiency of the estimator described here, we compare it to the ML estimator. A convenient way of doing this is to look at the "simulation residual", i.e. the difference between the score and the simulated score. First, (3.8) with Z_n replaced by $\bar{Z}_n$ is compared with (3.12). Then the following residual is obtained:

$$\frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R}\left[\bar{Z}_n(P_n - k_{nr}) - \left\{\bar{m}_{nr} - (1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \vartheta}\right\}\right] \tag{3.14}$$

The dummy variable can be rewritten as

$$\begin{aligned} d_n &= P_n + \nu_n \\ \text{with } E(\nu_n) &= 0 \\ \text{and } Var(\nu_n) &= P_n(1-P_n) \end{aligned} \tag{3.15}$$

Inserting this in the residual gives:

$$\frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R}\left[\bar{Z}_n(P_n - k_{nr}) - \left\{\bar{m}_{nr} - \frac{\partial \bar{P}_n}{\partial \vartheta}\right\}\right] - \sum_{n=1}^{N}\nu_n \frac{\partial \ln \bar{P}_N}{\partial \vartheta} \tag{3.16}$$

The variance of the first term of (3.16) can be reduced by increasing the number of drawings R. Suppose that $Var[\bar{Z}_n(P_n - k_{nr}) - \{\bar{m}_{nr} - \frac{\partial \bar{P}_n}{\partial \vartheta}\}] = \Xi_n$, conditional on the instruments $\bar{Z}_n$. This variance does not depend on R because the drawings are identical and independent. Then the variance of the first term is $\frac{1}{R}\sum_{n=1}^{N}\Xi_n$. With fixed N, increasing R to infinity results in reducing this variance to zero. The second term is the error which is caused by the fact that $(1-d_n)\frac{\partial \ln \bar{P}_n}{\partial \vartheta}$ is simulated by a simulator for $\frac{\partial \bar{P}_n}{\partial \vartheta}$. The expectation of this term equals zero, whereas the variance equals $\sum_{n=1}^{N}\bar{P}_n P_n \frac{\partial \ln \bar{P}_n}{\partial \vartheta}\frac{\partial \ln \bar{P}_n}{\partial \vartheta'}$. This term of the simulation residual cannot be influenced by the number of drawings. Therefore, this term leads to inefficiency, also for large R.

To compare the efficiency of the method of simulated scores estimator to the maximum likelihood estimator, also the term involving the difference between $\bar{Z}_n$ and Z_n has to be taken into account. The simulation residual then becomes:

$$RES = \frac{1}{R}\sum_{n=1}^{N}\sum_{r=1}^{R}\left[Z_n(P_n - k_{nr}) - \left\{\bar{m}_{nr} - \frac{\partial \bar{P}_n}{\partial \vartheta}\right\}\right] + \sum_{n=1}^{N}(\bar{Z}_n - Z_n)(d_n - k_n) - \sum_{n=1}^{N}\nu_n \frac{\partial \ln \bar{P}_n}{\partial \vartheta} \tag{3.17}$$

Compared to (3.16) we now have an additional term involving the difference $\bar{Z}_n - Z_n$. If we base $\bar{Z}_n$ on (3.9), with some consistent estimate of ϑ and simulators of $\frac{\partial \bar{P}_n}{\partial \vartheta}$ and P_n based on R_Z replications, then for R_Z going to infinity the difference disappears. Note that during the optimization process the matrix of instruments $\bar{Z}$ is fixed. This implies that this matrix needs to be initialized only once at the beginning of the estimation procedure. Therefore it is computationally feasible to calculate the instrument matrix using a large number of drawings R_Z, which need not be equal to the number of drawings R that is used in the calculation of the remaining part of the simulated score.

3.1 Asymptotic distribution of the estimator

The simulated score vector satisfies the property that its expectation, evaluated at the true parameter vector ϑ_0, equals zero. It is intuitively clear that if we solve the moment equations, defined by the simulated score, with respect to the parameter vector, the resulting parameter vector $\hat{\vartheta}_R$, at which the simulated score is zero, will converge to the true parameter value ϑ_0, or, equivalently, $\hat{\vartheta}_R$ will be a consistent estimator of ϑ_0. If smooth simulators were used, standard asymptotic theory could be applied to derive the consistency and asymptotic normality of the estimator. Pakes and Pollard (1989) derive conditions for consistency and asymptotic normality that do not rely on smoothness assumptions. The first set of conditions concerns the moment vector (3.11). Let $G(\vartheta)$ denote the unconditional expectation of the term in square brackets in (3.11). $G(\vartheta)$ is a set of population moments. Note that $G(\vartheta_0) = 0$. The identifiability condition requires that $\inf_{\|\vartheta-\vartheta_0\|>\delta} \| G(\vartheta) \| > 0 \forall \delta > 0$. Furthermore, $G(\vartheta)$ is required to have a non-singular derivative matrix at ϑ_0.

$\frac{1}{N}K_R(\vartheta)$ is the empirical counterpart of $G(\vartheta)$. In their proof of consistency and asymptotic normality, Pakes and Pollard use the fact that $\frac{1}{N}K_R(\vartheta)$ can be written as the expectation with respect to an empirical distribution function, the shape of which depends on the particular simulator that is employed, which should converge to its population counterpart $G(\vartheta)$, making use of the independence-across-observations assumption. Consequently, no smoothness assumptions are required. A condition which has to be satisfied in case of using a frequency simulator is that the probability of being at a tie (i.e. $d = 1$ <u>and</u> $d = 0$) has to be zero at ϑ_0. This condition is clearly satisfied here. Finally, Pakes and Pollard, as opposed to McFadden, allow the region which determines whether $d = 1$ or $d = 0$ to be a non-smooth function of the parameters as well. For our application non-smoothness of this region is not required. Smoothness of this region, together with the above conditions, is sufficient for consistency and asymptotic normality:

$$\frac{1}{\sqrt{N}}K_R(\vartheta_0) \widetilde{\text{asy}} N(0, V_{R0}), \tag{3.18}$$

with V_{R0} some positive definite symmetric matrix, and

$$\sqrt{N}(\hat{\vartheta}_R - \vartheta_0) \quad \widetilde{\text{asy}} \quad N(0, \Gamma^{-1}V_{R0}(\Gamma')^{-1}) \tag{3.19}$$

$$\text{where } \Gamma' \quad = \quad plim\frac{1}{N}\frac{\partial(\frac{\partial L(\vartheta_0)}{\partial \vartheta'})}{\partial \vartheta} \tag{3.20}$$

Using the expression of the asymptotic covariance matrix and the results of the analysis of the simulation residuals, it is possible to analyse the efficiency of the estimators by comparing the asymptotic covariance matrices of the simulation estimators with the asymptotic covariance matrix of the maximum likelihood estimator. It is a well known result that

$$\sqrt{N}(\hat{\vartheta}_{ML} - \vartheta_0) \quad \widetilde{\text{asy}} \quad N(0, \Omega_{ML}) \tag{3.21}$$

$$\text{where } \Omega_{ML} \quad = \quad B^{-1} \tag{3.22}$$

$$B = -plim\frac{1}{N}\frac{\partial^2 L(\vartheta_0)}{\partial \vartheta \partial \vartheta'} \tag{3.23}$$

To make clear the relation with the asymptotic covariance matrix of the simulation estimators we rewrite Ω_{ML} as

$$\Omega_{ML} = \Gamma_{ML}^{-1}V_{ML}(\Gamma'_{ML})^{-1} \tag{3.24}$$

where

$$\Gamma'_{ML} = plim\frac{1}{N}\frac{\partial\left(\frac{\partial L(\vartheta_0)}{\partial \vartheta'}\right)}{\partial \vartheta} = -B \tag{3.25}$$

which is the equivalent of (3.22), and

$$V_{ML} = plim \frac{1}{N} \sum_{n=1}^{N} \frac{\partial L_n(\vartheta_0)}{\partial \vartheta} \frac{\partial L_n(\vartheta_0)}{\partial \vartheta'} \tag{3.26}$$

To examine the efficiency of the estimator we first need to establish the relation between Γ'_{ML} and Γ'. It is readily established that

$$\Gamma' - \Gamma'_{ML} =$$
$$-plim \tfrac{1}{N} \textstyle\sum_{n=1}^{N} \left[(\bar{Z}_n - Z_n) \tfrac{\partial P_n}{\partial \theta} + \tfrac{\partial Z'_n}{\partial \theta} (d_n - P_n) + \nu_n \tfrac{\partial \left(\frac{\partial \ln \bar{P}_n}{\partial \theta'} \right)}{\partial \theta} + \tfrac{1}{\bar{P}} \tfrac{\partial \bar{P}_n}{\partial \theta} \tfrac{\partial \bar{P}_n}{\partial \theta'} \right] \tag{3.27}$$

from which only the first three terms equal zero if the instruments are constructed according to (3.9) with the number of drawings per individual tending to infinity. From the analysis of the simulation residuals it becomes clear that if the matrix of instruments is constructed according to (3.9) with drawings tending to infinity, and if the response probabilities and their derivatives are simulated with R tending to infinity as well, the asymptotic variance of the score of the likelihood function, evaluated in a consistent estimator is exceeded by X, where

$$X = plim \left(\frac{1}{N} \sum_{n=1}^{N} P_n \bar{P}_n \frac{\partial \ln \bar{P}_n}{\partial \vartheta} \frac{\partial \ln \bar{P}_n}{\partial \vartheta'} \right) \tag{3.28}$$

To estimate the covariance matrix we calculate

$$\hat{\Omega}_R = \hat{\Gamma}^{-1} \hat{V}_R (\hat{\Gamma}')^{-1} \tag{3.29}$$

with

$$\hat{\Gamma}' = \frac{1}{N} \frac{\partial \left(\frac{\partial L(\hat{\vartheta}_R)}{\partial \vartheta'} \right)}{\partial \vartheta} \tag{3.30}$$

$$\hat{V}_R = \frac{1}{N} \sum_{n=1}^{N} K_{nR}(\hat{\vartheta}_R) K_{nR}(\hat{\vartheta}_R)' \tag{3.31}$$

where the index n indicates the n-th component of the simulated score. Expression (3.31) can be calculated by simulation.

4 Stochastic specification

Recall (2.12):

$$\theta_n = \theta_0 + x'_n \omega + v_n \tag{4.1}$$

We have seen above that for the specification of the utility function adopted here, indifference curves will be convex whenever the utility function is defined. The utility function is defined whenever (2.10) holds true. We want (2.10) to hold true for all data points. To indicate this, we add subscripts and write:

$$c_n < -\theta_n + \frac{\gamma}{2\beta^2} + \frac{(h_n - \delta)^2}{2\gamma} := f_n(h_n) \tag{4.2}$$

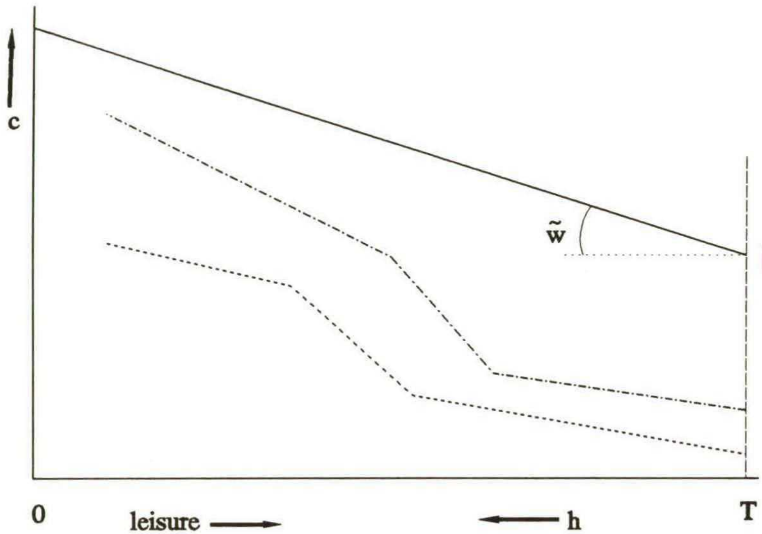

Figure 3: An encompassing budget set

To ensure that the direct utility function is properly defined for every individual in the sample, a practical procedure is the following one. Let $\tilde{w}$ and $\tilde{\mu}$ be the wage rate and nonlabor income which imply a linear budget constraint such that all observed budget sets are contained in it. We call this an *encompassing budget set*. See Fig. 3 for an illustration.

If we restrict the range of θ_n such that inequality (4.2) holds for all values of c_n and h_n in this encompassing budget set, then we know that indifference curves are convex at all data points. To achieve this we have to restrict the range of θ_n such that the function $f_n(.)$ is either tangent to the encompassing budget constraint or is outside the encompassing budget set. A tangency point is found for $h_n = \delta + \tilde{w}\gamma$ and

$$\theta_n = -\tilde{\mu} - \delta\tilde{w} - \frac{1}{2}\gamma\tilde{w}^2 + \frac{\gamma}{2\beta^2} \tag{4.3}$$

Thus, in view of (4.2) the inequality constraint on θ_n has to be

$$\theta_n < -\tilde{\mu} - \delta\tilde{w} - \frac{1}{2}\gamma\tilde{w}^2 + \frac{\gamma}{2\beta^2} \tag{4.4}$$

To guarantee that this inequality on θ_n holds for all observations we proceed as follows. Let the error term v_n be defined on $(-\infty, 0)$ then we impose the restriction:

$$\theta_0 < -x_n'\omega - \tilde{\mu} - \delta\tilde{w} - \frac{1}{2}\gamma\tilde{w}^2 + \frac{\gamma}{2\beta^2} \tag{4.5}$$

for all n. For the random preference term v_n we will actually assume that it follows a negative Γ distribution, defined on $(-\infty, 0)$. A similar procedure, in a somewhat different context, was followed by Kapteyn, Kooreman, and Van Soest (1990).

For non-participating individuals wages are not known and have to be integrated out. To ensure coherency of the model, the support of the wage distribution has to be restricted so that for all wages the implied budget set is contained within the encompassing budget set. This is achieved by restricting the support of the wage distribution to $[0, \tilde{w}]$. A convenient choice of distribution for u_n, which restricts the range of w_n to $[0, \tilde{w}]$, is to define a random variable λ_n following a lognormal distribution with log-mean m_n and log-variance τ^2, and to define

$$u_n := \log\{\psi_n \lambda_n / [1 + \lambda_n]\} \tag{4.6}$$

where

$$\psi_n := \tilde{w} / [\exp(w(x_n, \eta))] \tag{4.7}$$

It seems reasonable to require that the median of $\exp(u_n)$ equals one. This holds true if we specify

$$m_n = -\log(\psi_n - 1) \tag{4.8}$$

Thus the number of free parameters is equal to the case where the u_n were assumed normal with mean zero.[9]

5 Outline of simulation

In this section a description is given of the construction of the simulators. The technical details are presented in the appendix. Most of the simulation can be understood by reconsidering Fig. 2. For convenience, we repeat the basic features in Fig. 4 and add some notation. Figure 4 presents an example of a non-linear budget constraint and a non-convexity where the non-linearities arise from a tax system with two brackets and from the welfare and social security system. The budget constraint has three segments. On the first segment, on the right hand side of the figure, the individual works a positive number of hours, whereas at the same time he receives an unemployment benefit, say. Of each additional guilder of labour income the individual looses, say, $\alpha\%$ of the social security benefit, until, eventually, at hours H_0, nothing is left of the benefit. This results in a net wage rate on the first segment of w_0. On the second segment, between H_0 and H_1, no more benefits are received and therefore the net wage rate rises to w_1. At hours H_1 the next tax bracket is reached, which causes the net wage rate to fall to w_2 on the third segment. Now let h_n^j denote the optimal labour supply of individual n at a linear budget constraint with slope w_{nj} and intercept μ_{nj}, $j = 0, 1, ..., m$, and denote optimal labour supply by $\bar{h}_n$.

$$\bar{h}_n = h(w_n, \mu_n; \alpha, x_n' \omega + v_n) \tag{5.1}$$

[9] Notice that under normality the inequality (4.4) is violated with non-zero probability and hence the model would be incoherent.

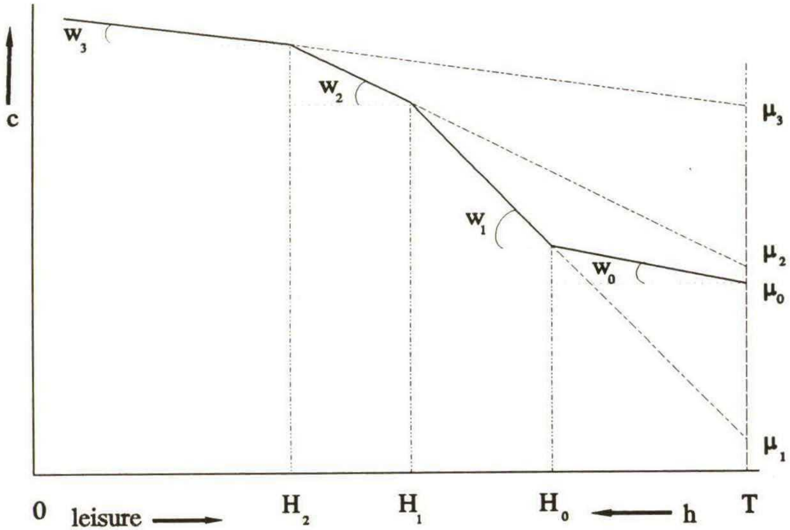

Figure 4: The determination of optimal hours

in which w_n is the before tax wage rate which implies that the function $h(.)$ includes the tax and the welfare system. Then the optimal labour supply $\bar{h}_n$, conditional on v_n and w_n, can be determined according to the following scheme:

$$\begin{array}{rlllllll}
h_{n,NC} & = 0 & if & & & h_n^0 & \leq & 0 \\
 & = h_n^0 & if & 0 & \leq & h_n^0 & \leq & H_{n0} \\
 & = H_{n0} & if & & & h_n^0 & > & H_{n0} \\
\tilde{h}_n & = H_{n0} & if & & & h_n^1 & < & H_{n0} \\
 & = h_n^j & if & H_{n,j-1} & < & h_n^j & \leq & H_{nj} \quad j = 1, ..., m-1 \\
 & = H_{nj} & if & h_n^j & > & H_{nj} & > & h_n^{j+1} \quad j = 1, ..., m-1 \\
 & = h_n^m & if & H_{n,m-1} & \leq & h_n^m & < & T \\
 & = T & if & & & h_n^m & \geq & T
\end{array} \tag{5.2}$$

$$\begin{array}{rll}
\bar{h}_n & = \tilde{h}_n & \text{if utility in } \tilde{h}_n \text{ exceeds utility in } h_{n,NC} \\
 & = h_{n,NC} & \text{otherwise}
\end{array}$$

$$h_n^j = \alpha_1 + \alpha_2 \mu_{nj} + \alpha_3 w_{nj} + \frac{1}{2}\alpha_4 w_{nj}^2 + x_n' \zeta + \alpha_2 v_n$$

The parameters α and ζ are obtained by reparametrization of β, γ, δ and ω in section 2. The precise form of this reparametrization is given in the appendix. The utility level has to be calculated using the direct utility function. Note that if the coherency restrictions are satisfied, the event $\bar{h}_n = H_{n0}$ will occur with probability zero.

Looking at variation in the gross wage rate w_n and in the unobservable taste component v_n, we can say that w_n determines the segments of the budget constraint by determining the slopes and the kink points, whereas v_n determines in which segment labour supply will be optimal given everything else. The distribution of measurement or optimization errors ϵ_n can be used to determine the distribution of h_n^*, defined in (2.15), conditional on the unobserved taste component v_n and on the gross wage rate w_n. The distribution of v_n can be used to integrate out the unobserved taste component, taking into account the decision rule on $\bar{h}_n$. Note that this involves comparing and integrating over utility functions. Finally, we multiply by the marginal density of the wage rate to obtain the joint density of h_n^* and w_n. This joint distribution can be used to determine the censored distribution of observed labour supply h_n and it will be clear that the expression for the probability that observed labour supply equals zero ($h_n = 0$) will be complicated. It will be difficult even to write down an analytic expression for the probability and the likelihood function as a whole. Therefore it will be impossible to use smooth simulators (see, e.g. McFadden,1989), because an analytic expression is needed in order to construct a smooth simulator. We will simulate the probability with a frequency simulator F_{nR}. The frequency simulator works as follows: Draw R times a wage rate w_{nr}^*, an unobserved taste variable v_{nr}^* and a measurement error ϵ_{nr}^* from their assumed distributions, calculate $\bar{h}_n$ and h_n^* and use the rules in (2.16) and (2.17) to determine h_n. The simulator becomes:

$$f_{nr} = \quad 1 \qquad \text{if } h_n > 0 \tag{5.3}$$

$$f_{nr} = \quad 0 \qquad \text{otherwise} \tag{5.4}$$

$$F_{nR} = \frac{1}{R}\sum_{r=1}^{R} f_{nr} \tag{5.5}$$

Since we also need the vector of derivatives of $P_n(\vartheta)$ we approximate this vector by a difference approximation of frequency simulators which is an unbiased simulator for the difference approximation of the probabilities. The evaluation of the contribution to the score vector of the working individuals involves the integration over random preferences. Two different methods to implement the integration are discussed in the appendix.

Finally, a suitable algorithm has to be chosen to minimize the objective function which can handle the discontinuities caused by the use of frequencies. Methods which make use of first derivatives turned out not to work and therefore we switched to the downhill simplex method of Nelder and Mead (1965) of which an overview is given in Press et al. (1986).

6 Results

In this section, the model is estimated using Monte Carlo data as well as real data. In the Monte Carlo experiment, the performance of the MSS method, outlined in section 3, is compared with some of the more conventional methods that have been mentioned in section 2. The conventional methods that we consider are the estimation of the model without random preferences and the estimation by instrumental variable methods. Furthermore, two variants of the MSS method are applied. The first variant estimates the parameters of the wage distribution separately by means of a reduced form wage-participation model. Then the labor supply parameters are estimated using predicted wages for non-participants. The second variant consists of estimating the wage and labor supply parameters simultaneously, thereby taking into account the stochastic nature of the budget constraint.

Subsequently the model is estimated for a sample of 849 married females, drawn from the Dutch population in 1985. In the Monte Carlo experiment the real data series of exogenous variables has been used in conjunction with a priori chosen parameter values to generate endogenous variables for each observation. The vector of taste shifters consists of the log family size variable (parameter ζ_1) and a dummy indicator taking the value one if the woman has children with age below 6, and zero if not (parameter ζ_2). In the Monte Carlo experiment, the variables in the wage equation are a constant term (parameter η_1), log-age (parameter η_2) and log-age squared (parameter η_3). To restrict the number of parameters in the Monte Carlo experiment we have omitted dummy indicators for the level of education. These dummies are included in the estimation on the real data. The parameters of the negative gamma distribution are γ_1 and γ_2. Using the distributional assumptions, random numbers have been generated which have been transformed to hours and wages using the true parameter values in the first column of table 6.1 and the decision rules in (2.16) and (2.17). The true parameter values are chosen such that the coherency restrictions are satisfied. Moreover we have tried to choose parameter values that generate distributions of observables similar to what we see in the sample. The values of some parameters are the result of experimentation with preliminary versions of the mode. Benefits are measured in guilders per week.

Both for the real data and the Monte Carlo data the budget constraint of each individual has been constructed on the basis of the Dutch tax code, also taking into account the welfare and social security system. In 1985 the tax system and the social security system were not well-integrated. They each have their own marginal tax rates and the social security system has ceilings for different sorts of payroll taxes. As a result of this the budget constraint may be quite complex with various kinks and with non-convexities.

Table 6.1 presents the a priori chosen parameter values which are used to generate the Monte Carlo data. Twenty Monte Carlo datasets have been generated using the completely specified model, which includes random preferences. [10]

[10]This small number of replications is due to the fairly heavy computational burden associated with estimation of the completely specified model

Table 6.1: True parameter vector	
Labor supply equation	
α_1 (Const.)	14.14
α_2 (non-labor income)	-0.04692
α_3 (wage)	10.690
α_4 (0.5× square of wage)	-0.260
ζ_1 (log(family size))	-24.002
ζ_2 (d. children with age < 6)	-13.903
Measurement error in hours	
σ_ϵ	19.31
Wage equation	
η_1 (const)	1.4
η_2 (log(age/17))	2.75
η_3 (square of log(age/17))	-2.2
Heterogeneity in wages	
τ	0.981
Random preferences	
γ_1	3.0
γ_2	4.0

6.1 Monte Carlo, no random preferences, predicted budget constraints

The first estimation method is the estimation of a simplified model which neglects random preferences and ignores random variation of budget constraints for non- participants. In the absence of random preferences the labor supply function (2.14) becomes

$$\bar{h}_n = h(w_n, \mu_n; \alpha, x_n'\omega) \tag{6.1}$$

Optimal labor supply is determined according to scheme (5.2) with $v_n = 0$. The only source of randomness now is measurement error ϵ and the participation rule in this model is:

$$\begin{aligned} h_n^* &= \bar{h}_n + \epsilon_n \\ h &= 0 \quad \text{if} \quad h_n^* \leq 0 \\ &= h_n^* \quad \text{if} \quad h_n^* > 0 \end{aligned} \tag{6.2}$$

The coherency constraint (4.5) remains the same.

For the non-participants one single budget-constraint is used, i.e. variation in the budget constraint due to variation in wages is ignored. This means that the wages for non-participants are predicted from the systematic part of the wage equation. For participants, the distributional assumptions (4.6), (4.7) and (4.8) are maintained. To avoid selection bias, the wage equation will be estimated for participants jointly with a selectivity equation of the form

$$y_n^* = \kappa' z_n + e_n \tag{6.3}$$

which can be interpreted as an approximate reduced form of

$$h_n^* = h(\exp(\eta' x_n + u_n), \mu_n; \alpha, x_n'\omega + v_n) + \epsilon \tag{6.4}$$

It is clear that the vector of variables z_n should contain all the variables included in the wage equation, as well as the variables appearing in the labor supply function $h(.)$. The joint wage-participation model is

$$\begin{array}{rcl} \ln w_n & = & \eta' x_n + u_n \\ y_n^* & = & \kappa' z_n + e_n \\ y_n^* & > & 0 \text{ if working } (w_n \text{ observed}) \\ & \leq & 0 \text{ if not } (w_n \text{ unobserved}) \end{array} \tag{6.5}$$

with

$$\begin{pmatrix} \ln \lambda_n \\ e_n \end{pmatrix} \sim N\left(\begin{pmatrix} m_n \\ 0 \end{pmatrix}, \Sigma \right) \tag{6.6}$$

and

$$\Sigma = \begin{pmatrix} \tau^2 & \sigma_{ue} \\ \sigma_{ue} & 1 \end{pmatrix} \tag{6.7}$$

where λ_n and m_n have been defined in (4.6) and (4.8). The variance of e_n has been normalized to one.

Once the model (6.5)-(6.7) has been estimated, (6.1) is estimated with predicted wages for non-participants. The predicted wages come from the systematic part of the wage equation.

Table 6.2 shows the Monte Carlo results for the estimation of this reduced form wage-participation model. The means in column 2 refer to the averages of estimates over 20 replications. Column three shows the standard deviations of the estimates over the replications and column four presents the average of the estimated asymptotic standard errors. Relative errors are given in the final column. They are defined by $|\bar{\theta} - \theta_0|/|\theta_0|$, where $\bar{\theta}$ is the mean in column 2 and θ_0 is the true parameter value. The mean of the parameter estimates $\hat{\eta}_2$ and $\hat{\eta}_3$, which correspond to the age variables in the wage equation are somewhat higher in absolute value than the true parameter values. The same holds for the variance τ. The parameters κ_j of the participation equation all are significant.

The Monte Carlo results for the labor supply parameters are given in table 6.3. The mean estimates of the constant term, α_2, ζ_1 and ζ_2 are all larger in absolute value than their true values, but their sign is estimated correctly. The estimate of α_3, the parameter of the linear wage term in the labor supply function, is close to its true value. However, the mean estimate of α_4, which corresponds to the quadratic wage term in the labor supply function is close to zero as compared with the true value. In fact the mean estimate of α_4 is about five standard deviations below the true value. The standard deviations and the mean SE are fairly similar. The variance σ_ϵ is higher than the true value, which is due to the neglect of random preferences.

6.2 Monte Carlo, IV, participants only

The next method we consider is the instrumental variables method. Now the non-convex piecewise linear budget constraint is linearized. Only participants are taken into consideration. We look at the observed value of labor supply, h. We check between which kinkpoints of the budget constraints this value is. Suppose that $H_{j-1} < h < H_j$: Then we are on the j-th segment and the budget constraint is linearized by the linear budget constraint with slope w_j and intercept μ_j that correspond to segment j. The equation we might want to estimate is:

$$h = \alpha_1 + \alpha_2\mu_j + \alpha_3 w_j + \frac{1}{2}\alpha_4 w_j^2 + x'\zeta + \epsilon \tag{6.8}$$

in which $w_j = (1 - \tau_j)w$, with w the gross wage rate and τ_j the marginal tax rate of segment j.

As is well known, there are two reasons why this equation cannot be simply estimated by OLS. The first is the presence of correlation between (μ_j, w_j) and the error term, and the second is the selectivity problem.

There are two causes for correlation between (μ_j, w_j) and the error term ϵ. The first is that the value of h determines which segment of the budget constraint is the appropriate segment, and consequently it determines which pair (μ_j, w_j) is chosen. Secondly, optimal labor supply need not coincide with observed labor supply due to measurement error. As the choice of the segment is determined by observed labor supply, instead of optimal labor supply, the wrong segment may be chosen. As a consequence, (μ_j, w_j) will be subject to measurement error as well, and their measurement errors are correlated with the measurement error of labor supply. The fact that we do not observe individuals at kink points can also be explained by measurement error. Instrumental variables for the intercept μ_j, the slope w_j and its square w_j^2 will have to be used. Obvious candidates are non-labor income μ, the gross wage rate w and its square, and individual characteristics that appear in the wage equation. It has to be assumed that the gross wage rate and ϵ are uncorrelated.

As we restrict ourselves to participants, a selectivity problem arises. We solve this problem by applying the standard Heckman correction: In (6.5) a reduced form wage-participation model has been presented. Suppose that the error term of the gross wage rate, u, is uncorrelated with the error ϵ of the labor supply equation. Next, make the standard assumption (false in this case) that the error term of the participation equation, e, and ϵ are jointly normally distributed. Then the expectation of ϵ, conditional on participation, $y^* > 0$, can be derived:

$$\lambda = \frac{\phi(-\kappa' z)}{1 - \Phi(-\kappa' z)} \tag{6.9}$$

in which $\phi(.)$ is the standard normal density function and $\Phi(.)$ the standard normal distribution function.[11] The estimate $\hat{\kappa}$, obtained from estimating the reduced form participation model can be used as a value for κ. To correct for selectivity, $\hat{\lambda}$ is added to the labor supply equation, where $\hat{\lambda}$ is equal to λ with κ replaced by $\hat{\kappa}$. The final estimation equation becomes:

$$h = \alpha_1 + \alpha_2\mu_j + \alpha_3 w_j + \frac{1}{2}\alpha_4 w_j^2 + x'\zeta + \sigma_{\epsilon e}\hat{\lambda} + \epsilon \tag{6.10}$$

[11] Strictly speaking we only assume (6.9), which is weaker than normality of e and ϵ

The model has been estimated on the Monte Carlo data, using two sets of instrumental variables. The *extended set* of instrumental variables contains, apart from the constant term, the correction term and the vector of characteristics x that already appears in the labor supply function, non-labor income μ, the gross wage rate w and its square and the age variables that also appear in the wage equation. These age variables are excluded from the *restricted set* of instrumental variables. Tables 6.4 and 6.5 present the Monte Carlo results obtained with the full set and the restricted set of instrumental variables respectively. There is not much difference between the Monte Carlo results with the different sets of instrumental variables. The signs of the estimates are correct, and there is significant evidence for the backward bending labor supply curve. The parameter α_3 of the linear wage term is a bit underestimated, whereas parameter α_4 of the quadratic wage term is over-estimated. The standard deviations of α_2, ζ_1 and ζ_2 are sizeable and the estimates are lower than the corresponding true values.

In the estimation with the instrumental variable methods, no coherency constraints are imposed and therefore these constraints may not be satisfied for all individuals. For the estimates obtained with the restricted set of instrumental variables, 6.3% of the individuals does not satisfy the coherency constraint. For the extended set this percentage is 8.3.

Comparing tables 6.4 and 6.5 with table 6.3, it appears that the instrumental variables method performs somewhat worse on average than the estimation of labor supply with neglect of random preferences, as carried out in the previous subsection, except for the fact that the instrumental variables method does manage to reproduce the backward bending labor supply curve.

6.3 Monte Carlo, MSS

We now consider the estimation by means of the method of simulated scores. Random preferences, as well as the tax and social security system are properly accounted for. Two variants can be distinguished. In the first variant we use predicted wages for non-working individuals. The predictors are obtained from the reduced form wage-participation model (6.5). Stochastic variation in the budget constraint due to stochastic variation in the wages is ignored here. The second method consists of estimating parameters of the labor supply model and the wage equation jointly.

In table 6.6 the Monte Carlo results of the variant with predicted wages are given. For the optimization of the objective function the downhill simplex method has been used. The results of the Monte Carlo study of the model without random preferences have been used in the construction of an initial starting simplex. The number of drawings to construct the simulator is equal to 10. The matrix of instruments has been constructed on basis of the true parameter values, using $R_Z = 800$ drawings. (Recall that the matrix of instruments has to be calculated only once at the beginning of the optimization procedure.) The consequences of constructing the matrix of instruments with estimated values in the context of a two step estimation procedure have been studied in Bloemen and Kapteyn (1993). It was found that the means of the estimated values do not change much by employing the two step procedure, but the standard deviations of the estimates are higher than in the case in which the matrix of instruments is calculated based on the true parameter vector.

The present method of estimation clearly outperforms the previous two estimation methods. Although the parameter of the quadratic wage term in the labor supply function is still underestimated, it is closer to its true value as well, and it is significantly different from zero at the 10% level.

Table 6.7 presents the Monte Carlo results of jointly estimating the labor supply model and the wage equation. Again, $R_Z = 800$ drawings were used to contruct the matrix of instruments and $R = 10$ drawings were used to construct the simulators. The estimates of the labor supply parameters are not very different from those obtained with the nonstochastic budget constraint in table 6.6. There is, however, a slight improvement in the estimation of the parameter of the quadratic wage term, which is closer to its true value than for any of the previous methods of estimation. Comparing the parameters of the wage distribution with the parameters obtained with the reduced form wage participation model in table 6.2, we see that there is an improvement in all but one of the parameters estimates.

In conclusion we may say that the Monte Carlo results show that the method of simulated scores with a nonstochastic budget constraint already yields fairly reasonable results, as compared to approximate methods like the instrumental variables method or leaving out random preferences. All of the methods of estimation seem to have problems in properly estimating the parameter of the quadratic wage term in the labor supply function. Leaving out random preferences severely underestimates the parameter of the quadratic wage term, whereas the instrumental variables method overestimates this parameter. In the Monte Carlo study the joint estimation of the labor supply function and the wage equation gives the best results with respect to the quadratic wage effect.

6.4 Estimation, no random preferences, predicted budget constraints

The model without random preferences is estimated using the 1985 OSA data, which includes 849 married female individuals of which 331 have a paid job. First the wage-participation model is estimated and the estimates are presented in table 6.8. Apart from the age variables, four education dummies have been included in the wage equation and consequently also in the participation equation. Educl is a dummy variable for the lowest level of education. The highest level of education is taken as the reference category. The four education dummies in the wage equation are negative and significant and they are increasing with the level of education, as they should. The age-earnings profile reaches its maximum at the age of 36. The dummy for the number of children with age below 6 and log family size have a significant negative effect on participation. A higher level of education tends to reduce the probability of non-participation. The probability of participation rises with age until the age of 29 after which it decreases. The covariance σ_{ue} between wages and participation is insignificant.

The wage estimates are used to predict wages for the non-participants, after which the labor supply model without random preferences is estimated.

The parameter estimates are given in table 6.9. Non-labor income has a small but significant (at the 10% level) negative effect on labor supply. The parameter estimate of

α_3, which corresponds to the linear wage term in the labor supply equation is positive and significant at the 10% level. The quadratic wage term does not seem to play a significant role in the labor supply function, so the present estimates do not provide evidence for a backward bending labor supply curve. Both the presence of children with age below 6 and an increase in log family size have a significant negative effect on labor supply. The age variables turn out to be insignificant.

6.5 Estimation, IV, participants only

Tables 6.10 and 6.11 present the empirical results, obtained with the instrumental variables method. Again two sets of instrumental variables are used. The *restricted set* contains the constant term, the dummy for the presence of children with age below 6, log family size, the correction term and the age variables. The *full set* contains, in addition to the variables that have been included in the restricted set, the education dummies which also appear in the wage equation.

There is not much difference between the IV estimates obtained with the restricted set and the IV estimates obtained with the full set of instrumental variables. The difference with the estimates in table 6.9, obtained by the model without random preferences, are considerable. Non-labor income has a larger impact on labor supply according to the IV estimates. Remarkably the IV estimates provide evidence in favour of a backward bending labor supply curve, as opposed to the estimates in table 6.9 in which the parameter estimate of α_4 was insignificant. This is in line with the Monte Carlo results presented above. Also there IV generated by far the largest estimate (in absolute value) of the quadratic wage effect. According to the IV estimates, the dummy for presence of children with age below 6 has a positive, though insignificant, effect. Log family size still has a significantly negative effect, but its estimated impact on labor supply is much smaller than according to table 6.9. The age variables are insignificant for both types of estimators.

The percentage of individuals that does not satisfy the coherency constraint is 45 for the restricted set and 43 for the extended set of instrumental variables.

6.6 Estimation, MSS

Table 6.12 shows the estimation results with MSS and a non-stochastic budget constraint.[12] The standard errors of the individual characteristics in the labor supply function are high relative to the estimates.

Table 6.13 presents the estimates obtained with the method of simulated scores and a stochastic budget constraint. The estimate of α_3, the parameter of the linear wage term in the labor supply function, is larger than the estimates for this parameter that we obtained with the IV and no random preferences methods. Non-labor income has a larger impact as well. The standard errors of the parameters of the wage distribution are rather high. The same holds, to a lesser extent, for the the standard errors of the

[12]To be sure, throughout we assume that the budget constraint is non-stochastic from the viewpoint of the agent; however from the viewpoint of the econometrician the budget constraint is stochastic since we do not observe all sources of heterogeneity across individuals

parameters ζ_i. Comparing the estimates of the wage parameters obtained with MSS with the estimates obtained with the reduced form wage-participation model in table 6.8, it is clear that there are differences. In view of the large standard errors in table 6.13, this does not necessarily mean much. Again, the quadratic wage term in the labor supply function is not significant. The parameter estimate of γ_2 is of large magnitude and estimated imprecisely.

6.7 Wage and participation elasticities for different methods

Table 6.14 shows the wage and participation elasticities that are implied by the various methods of estimation. These have been calculated as "aggregate" elasticities in the sense that all wages in the sample have been raised by 5% and then hours and participation have been predicted for every individual in the sample. For the method of simulated scores with a stochastic budget constraint the wages for non-working individuals have been simulated using the estimates of the wage distribution in table 6.13. For the remaining estimation methods the predicted wages based on the estimates in table 6.8 have been used. For the simulation of hours and participation the scheme (5.2) with $v_n = 0$ has been used for the IV method and the method without random preferences, whereas simulated values for v_n have been inserted for the MSS methods. The wage elasticities range from 0.11 for MSS with a non-stochastic budget constraint up to 1.29 for the model without random preferences. The participation elasticities range from 0.064 to 0.99. The elasticities with the IV method have been calculated both including and excluding the individuals that do not satify their coherency constraint. The standard errors of the estimated elasticities are sizeable. Consequently, the differences between the elasticities are not significant.

Table 6.15 presents the wage and participation elasticities of the Monte Carlo data. Also for the Monte Carlo data the different estimation results produce different elasticities, although the variation is a bit less than for the real data. Note that the ranking of the elasticities by method of estimation coincides with the ranking of the empirical elasticities in table 6.14, except for MSS with a non-stochastic budget constraint, which exhibits much larger elasticities for the Monte Carlo data than for the real data. The standard errors of the elasticities of the Monte Carlo data are smaller than for the real data. For the Monte Carlo data, most of the differences in the elasticities are significant.

Altogether, it is clear that there are considerable differences in the empirical estimates obtained by the various methods of estimation. These differences in the estimates have implications for the wage and participation elasticities, which vary widely across different methods of estimation. The standard errors of the elasticities are sizeable. Apart from the IV method, none of the methods provide evidence in favour of a backward bending labor supply curve. In the Monte Carlo experiment we saw that IV tends to overestimate the quadratic wage term in the labor supply function, whereas the other methods had a tendency to underestimate.

Table 6.2 Monte Carlo: Wage-participation model

θ	true value	mean	SD	mean SE	rel. err.
Participation equation (6.3)					
κ_1 (const)	–	0.825	0.203	0.251	–
κ_2 (log(fam. size))	–	-0.536	0.0552	0.0864	–
κ_3 (d. child. $<$ 6)	–	-0.264	0.0604	0.0639	–
κ_4 (non-labor income)	–	-0.000819	0.0000940	0.0000742	–
κ_5 (log(age/17))	–	3.083	0.699	0.798	–
κ_6 (square of log(age/17))	–	-2.425	0.516	0.544	–
Wage equation (6.5)					
η_1 (const)	1.4	1.097	0.169	0.235	0.22
η_2 (log(age/17))	2.75	3.408	0.564	0.698	0.24
η_3 (square of log(age/17))	-2.2	-2.646	0.440	0.483	0.20
τ	0.981	1.145	0.0591	0.0388	0.17
σ_{ue}	–	1.136	0.0632	0.0406	–

Table 6.3 Monte Carlo: Labour supply model
No random preferences

θ	true value	mean	SD	mean SE	rel. err.
α_1 (const)	14.1	15.768	5.734	9.456	0.12
α_2 (non-labor income)	-0.047	-0.0631	0.00980	0.00648	0.34
α_3 (wage)	10.69	10.704	1.286	2.270	0.0014
α_4 (0.5 $\times$ square of wage)	-0.26	-0.0126	0.0477	0.115	0.95
ζ_1 (log(family size))	-24	-34.044	5.218	3.788	0.42
ζ_2 (d. children with age $<$ 6)	-13.9	-17.594	5.607	3.328	0.27
σ_ϵ	19.3	24.234	0.729	0.856	0.26

Table 6.4 Monte Carlo: Labour supply model
The Instrumental Variables method
Extended set of Instrumental Variables

θ	true value	mean	SD	mean SE	rel. err.
α_1 (const)	14.1	28.657	6.808	7.618	1.03
α_2 (non-labor income)	-0.047	-0.0284	0.0207	0.0209	0.40
α_3 (wage)	10.69	7.391	1.991	2.210	0.31
α_4 (0.5 $\times$ square of wage)	-0.26	-0.355	0.160	0.158	0.36
ζ_1 (log(family size))	-24	-11.236	7.151	5.354	0.53
ζ_2 (d. children with age $<$ 6)	-13.9	-5.868	3.481	3.905	0.59
σ_ϵ	19.3	27.352	2.907	—	0.42

Table 6.5 Monte Carlo: Labour supply model
The Instrumental Variables method
Restricted set of Instrumental Variables

θ	true value	mean	SD	mean SE	rel. err.
α_1 (const)	14.1	26.591	6.934	8.234	0.88
α_2 (non-labor income)	-0.047	-0.0351	0.0244	0.0232	0.25
α_3 (wage)	10.69	8.071	1.965	2.401	0.24
α_4 (0.5 $\times$ square of wage)	-0.26	-0.377	0.165	0.170	0.45
ζ_1 (log(family size))	-24	-12.069	7.826	5.677	0.50
ζ_2 (d. children with age < 6)	-13.9	-6.567	3.940	4.176	0.53
σ_ϵ	19.3	28.623	3.707	—	0.48

Table 6.6 Monte Carlo: Labour supply model
Method of Simulated Scores, R = 10
Non-stochastic budget constraint

θ	true value	mean	SD	SE	rel. err.
α_1 (const)	14.1	13.639	9.573	7.683	0.035
α_2 (non-labor income)	-0.047	-0.0504	0.0116	0.00335	0.075
α_3 (wage)	10.69	11.041	1.182	1.004	0.033
α_4 (0.5 $\times$ square of wage)	-0.26	-0.169	0.101	0.360	0.35
ζ_1 (log(family size))	-24	-25.693	9.071	7.351	0.070
ζ_2 (d. children with age < 6)	-13.9	-14.604	6.032	7.409	0.050
σ_ϵ	19.31	20.851	2.625	6.019	0.080
γ_1	3.0	3.025	1.319	2.239	0.0082
γ_2	4.0	4.111	0.883	2.561	0.11

Table 6.7 Monte Carlo:
Labour supply model and wage distribution
Method of Simulated Scores, R = 10
Stochastic budget constraint

θ	true value	mean	SD	SE	rel. err.
α_1 (const)	14.1	14.467	8.829	4.585	0.023
α_2 (non-labor income)	-0.047	-0.0517	0.0129	0.0138	0.10
α_3 (wage)	10.69	11.075	1.452	0.903	0.036
α_4 (0.5 $\times$ square of wage)	-0.26	-0.176	0.0967	0.210	0.32
ζ_1 (log(family size))	-24	-25.409	8.048	5.831	0.059
ζ_2 (d. children with age < 6)	-13.9	-14.750	4.112	4.107	0.061
σ_ϵ	19.31	23.966	5.475	2.481	0.24
γ_1	3.0	3.222	1.076	0.403	0.074
γ_2	4.0	4.711	2.176	1.061	0.18
The wage distribution					
η_1 (const)	1.4	1.444	0.365	0.197	0.044
η_2 (log(age/17))	2.75	3.395	0.547	0.439	0.23
η_3 (square of log(age/17))	-2.2	-3.149	0.812	0.334	0.43
τ	0.981	1.082	0.103	0.0195	0.10

Table 6.8 Estimates of the wage-participation model

θ	$\hat{\theta}$	SE
κ_1 (const)	2.922	0.531
κ_2 (log(fam. size))	-1.335	0.179
κ_3 (d. child. < 6)	-1.085	0.150
κ_4 (non-labor income)	-0.000340	0.000163
κ_5 (log(age/17))	3.136	1.108
κ_6 (square of log(age/17))	-2.970	0.739
κ_7 (educ1)	-1.819	0.435
κ_8 (educ2)	-1.983	0.435
κ_9 (educ3)	-1.521	0.426
κ_{10} (educ4)	-0.915	0.456
η_1 (const)	2.493	0.164
η_2 (log(age/17))	1.896	0.470
η_3 (square of log(age/17))	-1.277	0.335
η_4 (educ1)	-0.668	0.0941
η_5 (educ2)	-0.562	0.0847
η_6 (educ3)	-0.477	0.0564
η_7 (educ4)	-0.213	0.0609
τ	0.470	0.0134
σ_{ue}	0.0396	0.0743

Table 6.9 Estimates of the labor supply model
No random preferences

θ	$\hat{\theta}$	SE
α_1 (const)	15.049	11.819
α_2 (non-labor income)	-0.00720	0.00369
α_3 (wage)	3.200	1.745
α_4 (0.5 $\times$ square of wage)	-0.0315	0.127
ζ_1 (log(family size))	-30.995	4.563
ζ_2 (d. children with age < 6)	-22.787	3.645
ζ_3 (log(age/17))	2.726	26.15
ζ_4 (square of log(age/17))	-24.578	17.503
σ_ϵ	24.240	1.540

Table 6.10 Estimates of the labor supply model
Instrumental Variables: restricted set

θ	$\hat{\theta}$	SE
α_1 (const)	35.842	4.218
α_2 (non-labor income)	-0.0389	0.0237
α_3 (wage)	1.719	0.422
α_4 (0.5 × square of wage)	-0.0950	0.0252
ζ_1 (log(family size))	-6.422	2.948
ζ_2 (d. children with age < 6)	0.748	2.487
ζ_3 (log(age/17))	-18.154	11.957
ζ_4 (square of log(age/17))	4.684	8.963
σ_ϵ	9.722	—
Correction term	-3.380	2.543

Table 6.11 Estimates of the labor supply model
Instrumental Variables: Full set

θ	$\hat{\theta}$	SE
α_1 (const)	35.780	4.175
α_2 (non-labor income)	-0.0372	0.0235
α_3 (wage)	1.719	0.417
α_4 (0.5 × square of wage)	-0.0965	0.0247
ζ_1 (log(family size))	-6.307	2.938
ζ_2 (d. children with age < 6)	0.870	2.478
ζ_3 (log(age/17))	-18.280	11.928
ζ_4 (square of log(age/17))	4.770	8.941
σ_ϵ	9.699	—
Correction term	-3.498	2.535

Table 6.12 Estimates of the labor supply model
Method of Simulated Scores, R = 10
Non-stochastic budget constraint

θ	$\hat{\theta}$	SE
α_1 (const)	37.867	29.733
α_2 (non-labor income)	-0.0194	0.0140
α_3 (wage)	11.522	6.790
α_4 (0.5 × square of wage)	-0.00462	0.468
ζ_1 (log(family size))	5.902	4.800
ζ_2 (d. children with age < 6)	14.900	16.738
ζ_3 (log(age/17))	85.328	70.310
ζ_4 (square of log(age/17))	-21.219	12.290
σ_ϵ	16.547	4.966
γ_1	4.370	2.187
γ_2	30.871	22.872

Table 6.13 Estimates of the labor supply model and wage distribution Method of Simulated Scores, R = 10 Stochastic budget constraint

θ	$\hat{\theta}$	SE
α_1 (const)	31.145	23.677
α_2 (non-labor income)	-0.0480	0.0216
α_3 (wage)	10.285	3.729
α_4 (0.5 $\times$ square of wage)	-0.00243	0.318
ζ_1 (log(family size))	-26.251	19.765
ζ_2 (d. children with age $<$ 6)	-17.912	21.983
ζ_3 (log(age/17))	56.486	40.756
ζ_4 (square of log(age/17))	-36.518	29.024
σ_ϵ	12.429	2.877
γ_1	4.138	2.022
γ_2	23.877	24.309
The wage distribution		
η_1 (const)	0.630	1.101
η_2 (log(age/17))	2.666	1.832
η_3 (square of log(age/17))	-1.325	1.477
η_4 (educ1)	-0.935	0.998
η_5 (educ2)	-0.859	0.980
η_6 (educ3)	-0.719	0.825
η_7 (educ4)	-0.581	0.942
τ	0.653	0.0192

Table 6.14 Wage and participation elasticities

Method of estimation	wage elasticity	SE	participation elasticity	SE
No random preferences	1.29	1.17	0.99	1.24
IV, restricted set	0.23	0.70	0.15	0.74
IV, restricted set (excluding non-coherents)	0.21	0.69	0.16	0.73
IV, full set	0.19	0.70	0.12	0.67
IV, full set (excluding non-coherents)	0.19	0.70	0.13	0.68
MSS, non-stochastic budget constraint	0.11	0.05	0.064	0.04
MSS, stochastic budget constraint	0.39	0.43	0.29	0.21

Table 6.15 Monte Carlo: wage and participation elasticities

Method of estimation	wage elasticity	SE	participation elasticity	SE
No random preferences	1.05	0.11	0.73	0.08
IV, restricted set	0.33	0.21	0.12	0.19
IV, restricted set (excluding non-coherents)	0.36	0.19	0.16	0.19
IV, full set	0.26	0.18	0.09	0.15
IV, full set (excluding non-coherents)	0.30	0.16	0.08	0.15
MSS, non-stochastic budget constraint	0.95	0.20	0.61	0.26
MSS, stochastic budget constraint	0.69	0.11	0.40	0.11

7 Conclusions

Both the Monte Carlo results and the estimation results for real data show large variation of outcomes across estimation methods. For the Monte Carlo we know the true model and the results suggest that an incorrect treatment of the stochastic nature of the data may lead to large biases. Estimated wage and participation elasticities may easily be double or half the true elasticity if the wrong estimation method is applied.

For the real data, we do not know the true model, of course, but the huge variation in parameters and implied elasticities is disconcerting. The fact that the ordering of elasticities is by and large the same as for the Monte Carlo data is suggestive of the fact that also here an oversimplification of stochastic structure may be a cause of biased outcomes.

In itself the model considered in this paper is not claimed to be realistic. After all, it does not have any dynamic elements, no fixed costs of working, etc. The purpose of the paper has not been to build a fully realistic model of labor market behavior. Rather we have limited ourselves to a somewhat simplified environment in which agents are supposed to behave and then concentrated on a utility consistent specification behavior in that environment. Where our results seem to show the extreme importance of a correct (utility consistent) treatment of the stochastic structure of the model in such a case, we would anticipate even more relevance of such treatment in more complicated environments.

A Appendix. Simulation of the score

In this appendix the technical details of the simulation of the score will be worked out. The simulation of the score can be split up in two parts, i.e. the simulation of the participation probabilities and the simulation of the score of the continuous part of the likelihood function.

First, some notation is introduced. Let μ_{nj} denote the intercept of the j-th segment of the budget constraint, as indicated in figure 4, where $j = 1, ..., m$. The index $j = 0$ indicates the segment which introduces the non-convexity in the budget constraint. The slope of the j-th segment is denoted by w_{nj}, $w_{n0} < w_{n1}, w_{nj} > w_{n,j+1}, j = 1, ..., m-1$, and H_{nj} is the kink point between the j-th and $(j+1)$-th segment, $j = 0, ..., m-1$. If

$H_{n0} = 0$ we just have the model without social security system. If we allow for variation in the gross wage w_n, the slopes w_{nj} and the kink points H_{nj} will depend on w_n, so formally:

$$\begin{aligned} w_{nj} &= w_j(w_n) && \text{(A.1)} \\ H_{nj} &= H_j(w_n) && \text{(A.2)} \\ w_j'(w_n) &> 0 && \text{(A.3)} \\ H_j'(w_n) &< 0 && \text{(A.4)} \end{aligned}$$

As in section 5, we let h_n^j denote the optimal amount of labour supply if the budget constraint is linear with slope w_{nj} and intercept μ_{nj}, $j = 1, ..., m$. In our model:

$$h_n^j = h^\bullet(\alpha, w_{nj}, \mu_{nj}, \zeta, x_n) + \alpha_2 v_n \qquad \text{(A.5)}$$

where

$$h^\bullet(\alpha, w, \mu, \zeta, x) = \alpha_1 + \alpha_2\mu + \alpha_3 w + \frac{1}{2}\alpha_4 w^2 + x'\zeta \qquad \text{(A.6)}$$

Expressing α and ζ in terms of the original parameters gives:

$$\begin{aligned} \alpha_1 &= \delta + \beta\theta_0 && \text{(A.7)} \\ \alpha_2 &= \beta && \text{(A.8)} \\ \alpha_3 &= \gamma + \beta\delta && \text{(A.9)} \\ \alpha_4 &= \beta\gamma && \text{(A.10)} \\ \zeta &= \omega\beta && \text{(A.11)} \end{aligned}$$

Because $\beta < 0$ and $\gamma > 0$ we find that $\alpha_2 < 0$ and $\alpha_4 < 0$. Notation will be abbreviated by defining

$$h_{nj}^\bullet = h^\bullet(\alpha, w_{nj}, \mu_{nj}, \zeta, x_n) \qquad \text{(A.12)}$$

The unobserved taste parameter v_n is assumed to be distributed according to a negative gamma-distribution with parameters γ_1 and γ_2, i.i.d. over individuals, by which we mean that $-v_n$ has a gamma distribution. The probability density function of v_n is

$$g(v_n, \gamma_1, \gamma_2) = \frac{1}{\Gamma(\gamma_1)\gamma_2^{\gamma_1}}(-v_n)^{\gamma_1 - 1}\exp\left(\frac{v_n}{\gamma_2}\right), \gamma_1 > 0, \gamma_2 > 0, -\infty < v_n < 0 \qquad \text{(A.13)}$$

As pointed out in section 5, the distribution of the measurement errors is assumed to be normal with mean zero and variance σ_ϵ^2:

$$\phi(\epsilon_n, \sigma_\epsilon^2) = \frac{1}{\sqrt{2\pi}\sigma_\epsilon}\exp\left\{-\frac{1}{2\sigma_\epsilon^2}\epsilon_n^2\right\}, -\infty < \epsilon_n < \infty \qquad \text{(A.14)}$$

The wage distribution can be derived from assumptions (4.6), (4.7) and (4.8).

$$\begin{aligned} \kappa(w_n, \eta, \tau^2) &= \frac{1}{\sqrt{2\pi}\tau}\frac{\tilde{w}}{\tilde{w} - w_n}\frac{1}{w_n}\exp\left\{-\frac{1}{2\tau^2}\left[\log\left(\frac{w_n}{\tilde{w} - w_n}\right) - m_n\right]^2\right\} && \text{(A.15)} \\ m_n &= -\log\left(\frac{\tilde{w}}{\exp(\eta' x_n)} - 1\right), 0 < w_n < \tilde{w} && \text{(A.16)} \end{aligned}$$

It is a straightforward extension to incorporate correlation between wages and measurement errors. However, to restrict the introduction of notation, we abstain from it here.

The likelihood contribution of an individual will be formulated now. We make use of the scheme (5.2) for determining optimal labour supply and the participation rules described in (2.16) and (2.17). First note that the h_n^j in (5.2) all depend on the random preference parameter v_n, so given everything else, v_n determines in which segment of the budget constraint labour supply is optimal. Therefore, we have to determine which set of values of v_n coincides with which segment of the budget constraint. The following sets are defined:

$$\begin{aligned}
A_j &= \{v_n | H_{n,j-1} < h_n^j \leq H_{nj}\} && j = 0, ..., m \\
B_0 &= \{v_n | h_n^0 \leq 0\} \\
B_j &= \{v_n | h_n^j > H_{nj} > h_n^{j+1}\} && j = 1, ..., m-1 \\
B_m &= \{v_n | h_n^m \geq T\} \\
& H_{n,-1} = 0, H_{nm} = T \\
Q(h_{n,NC}, h_{nj}) &= \{v_n | U(h_{n,NC}, y_0(h_{n,NC})) < U(h_{nj}, y_j(h_{nj}))\} \\
Q^*(h_{n,NC}, h_{nj}) &= \{v_n | U(h_{n,NC}, y_0(h_{n,NC})) > U(h_{nj}, y_j(h_{nj}))\} \\
& \text{with } h_{n,NC} \text{ defined in (5.2) and} && y_j(h) = w_j h + \mu_j \\
R_j(h_{n,NC}, h_n^j) &= A_j \cap Q^*(h_{n,NC}, h_n^j) && j = 1, ..., m \\
S_j(h_{n,NC}) &= B_j \cap Q^*(h_{n,NC}, H_{nj}) && j = 1, ..., m \\
Z_{01} &= B_0 \cap \left\{ \left(\cup_{j=1}^m R_j(0, h_n^j)\right) \cup \left(\cup_{j=1}^m S_j(0)\right) \right\} \\
Z_{02} &= A_0 \cap \left\{ \left(\cup_{j=1}^m R_j(h_n^0, h_n^j)\right) \cup \left(\cup_{j=1}^m S_j(h_n^0)\right) \right\} \\
Z_{j1} &= B_j \cap Q(h_{n,NC}, H_{nj}) && j = 1, ..., m \\
Z_{j2} &= A_j \cap Q(h_{n,NC}, h_n^j) && j = 1, ..., m \\
Z^* &= \left(\cup_{j=1}^m Z_{j1}\right) \cup \left(\cup_{j=0}^m Z_{j2}\right)
\end{aligned} \tag{A.17}$$

Z_{01} is the set of v_n for which optimal labour supply is zero, Z_{02} is the set for which it is optimal to be on the first segment of the budget constraint, before the nonconvexity kink H_{n0}, Z_{j1} is the set for which optimal labour supply is equal to the j-th kink, $j = 1, ..., m$,

Z_{j2} is the set for which it is optimal to be on the j-th segment after the nonconvexity kink, $j = 1, ..., m$ and Z^* is the set for which optimal labour supply is positive.

We now determine the probability that observed labour supply is zero, conditional on the value of v_n. According to (2.17) there are two possibilities for observed labour supply to be zero. The first happens when optimal labour supply is zero. Then observed labour supply is equal to zero, irrespective of the value of measurement error. So if v_n is from the set for which optimal labour supply is zero, the probability that observed labour supply is zero, conditional on v_n, is equal to one. The second possibility for observed labour supply to be zero occurs when optimal labour supply is positive but optimal labour supply plus measurement error is negative. Summarizing, the probability that observed labour supply is zero, conditional on v_n becomes:

$$\begin{aligned} P(h_n = 0|v_n, w_n) &= 1 \quad \text{if} \quad v_n \in Z_{01} \\ &= \Phi\left(-\tfrac{H_{nj}}{\sigma_\epsilon}\right) \quad \text{if} \quad v_n \in Z_{j1}, j = 1, ..., m \\ &= \Phi\left(-\tfrac{h_n^j}{\sigma_\epsilon}\right) \quad \text{if} \quad v_n \in Z_{j2}, j = 0, ..., m \end{aligned} \tag{A.18}$$

in which $\Phi(.)$ is the standard normal distribution function.

The contribution of positive values of labour supply, conditional on v_n, is restricted to $v_n \in Z^*$ for which optimal labour supply is positive.

$$\begin{aligned} \chi(h_n|v_n, w_n) &= \phi(h_n - H_{nj}, \sigma_\epsilon^2) \quad \text{if} \quad v_n \in Z_{j1}, j = 1, ..., m \\ \chi(h_n|v_n, w_n) &= \phi(h_n - h_n^j, \sigma_\epsilon^2) \quad \text{if} \quad v_n \in Z_{j2}, j = 0, ..., m \end{aligned} \tag{A.19}$$

Having determined the density of observed labour supply, conditional on v_n, the unconditional contribution can be obtained by integrating over v_n.

$$\begin{aligned} P(h_n = 0|w_n) &= \textstyle\int_{Z^* \cup Z_{01}} P(h_n = 0|v, w_n) g(v, \gamma_1, \gamma_2) dv \quad \text{if } h_n = 0 \\ l(h_n|w_n) &= \textstyle\int_{Z^*} \chi(h_n|v, w_n) g(v, \gamma_1, \gamma_2) dv \quad \text{if } h_n > 0 \end{aligned} \tag{A.20}$$

or, making use of (A.18) and (A.19)

$$P(h_n = 0|w_n) =$$

$$\int_{Z_{01}} g(v, \gamma_1, \gamma_2) dv + \sum_{j=1}^{m} \int_{Z_{j1}} \Phi\left(-\frac{H_{nj}}{\sigma_\epsilon}\right) g(v, \gamma_1, \gamma_2) dv + \sum_{j=0}^{m} \int_{Z_{j2}} \Phi\left(-\frac{h_n^j}{\sigma_\epsilon}\right) g(v, \gamma_1, \gamma_2) dv \tag{A.21}$$

$$\text{if } h_n = 0$$

$$l(h_n|w_n) = \sum_{j=1}^{m} \int_{Z_{j1}} \phi(h_n - H_{nj}, \sigma_\epsilon^2) g(v, \gamma_1, \gamma_2) dv + \sum_{j=0}^{m} \int_{Z_{j2}} \phi(h_n - h_n^j, \sigma_\epsilon^2) g(v, \gamma_1, \gamma_2) dv \tag{A.22}$$

$$\text{if } h_n > 0$$

For an individual whose labour supply is zero, wages are not observed and they are integrated out. The final response probability becomes

$$\int_0^{\bar{w}} P(h_n = 0|w)\kappa(w, \eta, \tau^2)dw \tag{A.23}$$

The problem with the above defined sets is that the bounds of these sets are not known explicitly. The advantage of the frequency simulator in the context of an integral with bounds that are known implicitly only, is that it is possible to draw random numbers and then check in which region the simulated value of labour supply is.[13]

We now describe the construction of the frequency simulator. The first thing we need is drawings from the distributions of measurement errors, wages and random preferences. As measurement errors are normally distributed, a series of R random numbers can be drawn from the standard normal distribution which will be kept constant during the minimization process. These basic drawings can be transformed to drawings from the distribution of ϵ_n through multiplying by σ_ϵ. Any change in the drawings of ϵ_n is caused by a change in σ_ϵ.

To draw a series of gross wages we also start by drawing a series of R standard normal random variables $l^*_{nr}, r = 1, ..., R$, which are the constant basic drawings. These basic drawings can be transformed to drawings of the wage rate:

$$w^*_{nr} = \frac{\tilde{w}\exp(m_n + \tau l^*_{nr})}{1 + \exp(m_n + \tau l^*_{nr})}, r = 1, ..., R \tag{A.24}$$

The transformation is continuous in the parameters and therefore, keeping the basic drawings constant, a change in the drawings w^*_{nr} can only be caused by a change in the parameters.

The generation of random numbers from the negative gamma distribution is not that straightforward as the generation of random numbers from a normal distribution. The method commonly used for the generation of gamma random numbers is the acceptance-rejection method. Although this method is very useful for generating gamma random numbers if the parameters remain constant, the use of this method in the context of a minimization problem with changing parameters is not appropriate. A change in the parameters can cause discrete jumps in the drawings. The alternative would be to generate random numbers by means of the inversion method, see e.g. Devroye (1986). A major drawback of this method is that for every draw the negative gamma distribution function has to be inverted using numerical methods. Experiments with the inversion method have shown that the application of this method in the context of an estimation problem leads to an infeasibly high computational burden, even in rather simple problems. Therefore, the inversion method applied in estimation by simulation procedures

[13]Note that it is possible to simulate (A.23) by drawing wages and random preferences from their respective distributions, without drawing measurement error, then checking the region (Z_{j1} or Z_{j2}) and setting the contribution to the simulator equal to the conditional probability corresponding to this region, evaluated in the drawings. This results in a piecewise continuous simulator, which is a combination of a frequency simulator and a smooth simulator. The possibility to construct this type of simulator however, depends strongly on the model structure imposed in (2.15)-(2.17), which would change if dynamic elements, fixed cost of working, separation of the labour supply decision and participation decision, etc. were introduced. Therefore, the general applicable frequency simulator is employed here.

is only useful either if the functional form of the inverse of the distribution function is known, or if a good approximation for the inverse of the distribution function is available. A third possibility is to use importance sampling. In that procedure the random numbers are drawn from a different distribution with favourable characteristics and it is corrected for drawing from a different distribution by the use of a weight function. This is the procedure which we use here. We draw random numbers from the exponential distribution with parameter ρ:

$$\Lambda(\rho, v_n) = \rho \exp\{\rho v_n\}, -\infty < v_n < 0, \rho > 0 \tag{A.25}$$

Because this is not the "true" (assumed) distribution, the frequency simulator has to be weighted like in importance sampling. The weight function $k(v_n, \gamma_1, \gamma_2, \rho)$ is the ratio of the negative gamma density function and the negative exponential density function.

$$k(v_n, \gamma_1, \gamma_2, \rho) = \frac{g(v_n, \gamma_1, \gamma_2)}{\Lambda(\rho, v_n)} = \frac{1}{\Gamma(\gamma_1)\gamma_2^{\gamma_1}\rho}(-v_n)^{\gamma_1 - 1} \exp\left\{\left(\frac{1}{\gamma_2} - \rho\right) v_n\right\} \tag{A.26}$$

The fact that we draw from the exponential distribution instead of the gamma distribution increases the variance of the estimator. In the first place we have to choose the parameter ρ in such a way that the variance will be finite and second, the choice of ρ has to make the addition to the variance as small as possible. In the implementation a random number v from the negative exponential density in (A.25) is inserted in (A.26), so in calculating the mean and the variance of the weight function we do this with respect to the negative exponential density. By construction, the mean of the weight function is always equal to one. Note that if it is drawn from the true density the weight function is identically equal to one and as a consequence the variance of the weight function is equal to zero. Therefore, the larger is the deviation of the shape of the approximate density function from the true density function, the larger will be the variance, see e.g. Kloek and Van Dijk (1978). The expression for the variance is given by:

$$\begin{gathered} E[k(v, \gamma_1, \gamma_2, \rho)]^2 - 1 = \\ \int_{-\infty}^{0} \left[\frac{g(v, \gamma_1, \gamma_2)}{\Lambda(\rho, v)}\right]^2 \Lambda(\rho, v) dv - 1 = \\ \int_{-\infty}^{0} k(v, \gamma_1, \gamma_2, \rho) g(v, \gamma_1, \gamma_2) dv - 1 = \\ \frac{\Gamma(2\gamma_1 - 1)\left(\frac{\gamma_2}{2 - \rho\gamma_2}\right)^{2\gamma_1 - 1}}{[\Gamma(\gamma_1)]^2 \gamma_2^{2\gamma_1} \rho} \end{gathered} \tag{A.27}$$

in which

$$\gamma_1 > \frac{1}{2} \tag{A.28}$$

$$\rho < \frac{2}{\gamma_2} \tag{A.29}$$

This is the difference of the mean of the weight function with respect to the true density function $g(v, \gamma_1, \gamma_2)$ and the mean of the weight function with respect to $\Lambda(v, \rho)$ which is equal to one. (A.29) is the necessary condition for the variance to be finite. The smallest variance can be obtained by choosing ρ in such a way that the variance of the weight function is minimized. Solving the first order conditions and checking the second order conditions, it can be found that the variance is minimal for

$$\rho = \frac{1}{\gamma_1 \gamma_2} \tag{A.30}$$

Note that condition (A.29) is satisfied if condition (A.28) is satisfied.

Summarizing, the drawing procedure for v_n is as follows. Draw a series of R random numbers $\tilde{v}^*_{nr}$ from the exponential distribution with parameter ρ. These are our basic drawings. Transform the basic drawings to drawings v^*_{nr} from an exponential distribution with parameter $\frac{1}{\gamma_1\gamma_2}$ by multiplying the basic drawings by $\gamma_1\gamma_2$:

$$v^*_{nr} = \gamma_1 \gamma_2 \tilde{v}^*_{nr} \tag{A.31}$$

Note that this is a continuous transformation in γ_1 and γ_2. These are the final drawings which will be used in the simulation of the labour supply.

Having described the way of generating the required random numbers, we now turn to the simulation of the probability. Using the drawings $(\epsilon^*_{nr}, w^*_{nr}, v^*_{nr})$ the optimal labour supply $\bar{h}_{nr}$ and the observed labour supply h_{nr} can be simulated according to scheme (5.2) and the participation rules (2.16) and (2.17). Then the participation probability can be simulated by a frequency simulator like in (5.3)-(5.5) where (5.3) has to be weighted. The frequency simulator becomes:

$$
\begin{aligned}
f_{nr} &= k(v^*_{nr}, \gamma_1, \gamma_2, \rho) \quad \text{if } h_{nr} > 0 && \text{(A.32)}\\
f_{nr} &= 0 \quad \text{otherwise} && \text{(A.33)}\\
F_{nR} &= \tfrac{1}{R} \textstyle\sum_{r=1}^{R} f_{nr} && \text{(A.34)}
\end{aligned}
$$

By construction, this is an unbiased simulator for the participation probability.

The estimation method also requires a simulator of the derivatives of the probability with respect to the parameters. Let $F_{nR}(\theta)$ denote the frequency simulator in parameter vector θ. Then the derivative with respect to the k-th component of θ is simulated by a difference interval of frequency simulators:

$$\bar{m}_{nk}(\theta, \epsilon^*_R, v^*_R, w^*_R) = \frac{F_{nR}(\theta + \delta e_k) - F_{nR}(\theta)}{\delta} \tag{A.35}$$

where e_k is the k-th unit vector. Because $F_{nR}(\theta + \delta e_k)$ is an unbiased simulator of the participation probability in $\theta + \delta e_k$ and $F_{nR}(\theta)$ is an unbiased simulator of the participation probability in θ, (A.35) is an unbiased simulator of the difference interval of the participation probability. Because $F_{nR}(\theta)$ is discontinuous in the parametervector θ we have to choose δ large enough to ensure that the sum of the difference interval over all individuals and all drawings in (3.12) is not equal to zero. The larger is the number of drawings R, the smaller can be the value of δ. To construct the optimal matrix of instruments, which only has to be calculated once at the beginning of the optimization procedure, a large number of drawings can be used. In our empirical applications we used 800 drawings.

We now turn to the simulation of the continuous part of the score vector. First, we abstract from the problems that arise because the bounds are unknown, and from the problem that we cannot draw directly from the gamma distribution. Note that it is possible to simulate the integral appearing in the log-likelihood function unbiasedly. Draw random numbers v^*_{nr} from the density $g(v,\gamma_1,\gamma_2)$ restricted to the region Z^* for which optimal labour supply is positive, defined in (A.17).

$$v^*_{nr} \sim \frac{g(v,\gamma_1,\gamma_2)}{P(v \in Z^*)}, v^*_{nr} \in Z^* \tag{A.36}$$

An unbiased simulator for $l(h_n|w_n)$ is

$$\tilde{l}(h_n|w_n) = P(v \in Z^*)\frac{1}{R}\sum_{r=1}^{R}\chi(h_n|v^*_{nr}, w_n) \tag{A.37}$$

or, writing this out:

$$\tilde{l}(h_n|w_n) =$$

$$P(v \in Z^*)\frac{1}{R}\sum_{r=1}^{R}\left\{\sum_{j=0}^{m} I(v^*_{nr} \in Z_{j2})\phi(h_n - h^j_n, \sigma^2_\epsilon) + \sum_{j=1}^{m} I(v^*_{nr} \in Z_{j1})\phi(h_n - H_{nj}, \sigma^2_\epsilon)\right\} \tag{A.38}$$

in which h^j_n is computed on the basis of v^*_{nr} and I(.) the indicator function. A simulator for the score contribution then could be obtained by simulating numerator and denominator in

$$\frac{\partial \ln l(h_n|w_n)}{\partial \vartheta} \tag{A.39}$$

separately. This introduces a bias in the simulation of the score in the sense that the expectation evaluated in the true parameter vector will not be equal to zero. However, this simulator for the continuous part of the score contribution is piecewise continuous and therefore it does not have the unfavourable characteristics of a probability frequency simulator in the context of simulated maximum likelihood, see e.g. Lerman and Manski (1981).

An additional complication arises from the fact that the bounds of the region Z^* are unknown. Hence we have to draw from a different region $\tilde{Z}^*$ which contains the original region, i.e. $Z^* \subset \tilde{Z}^*$, and which approximates the original region as close as possible. Consider the region

$$\begin{aligned} \tilde{Z}^* &= \{v| -\infty < v < q(\alpha,\zeta,w_n,\mu_n)\} \\ q(\alpha,\zeta,w_n,\mu_n) &= -h^\bullet_{n1}/\alpha_2 \text{ if } -h^\bullet_{n1}/\alpha_2 < 0 \\ &= 0 \text{ if } -h^\bullet_{n1}/\alpha_2 > 0 \end{aligned} \tag{A.40}$$

$-h^\bullet_{n1}/\alpha_2$ is the value of v for which h^1_n is equal to zero. The region Z^* of positive optimal labour supply is contained in this region. The simulation procedure now becomes as follows: Draw a random number v^*_{nr} from the negative exponential distribution with parameter ρ:

$$v^*_{nr} \sim \rho\exp\{\rho v^*_{nr}\}, -\infty < v^*_{nr} < 0 \tag{A.41}$$

Define ν^*_{nr} as

$$\nu^*_{nr} = v^*_{nr} + q(\alpha, \zeta, w_n, \mu_n) \tag{A.42}$$

As a consquence, ν^*_{nr} is a truncated negative exponential variable:

$$\nu^*_{nr} \sim \frac{\rho \exp\{\rho \nu^*_{nr}\}}{P(\nu \in \tilde{Z}^*)}, \nu^*_{nr} \in \tilde{Z}^* \tag{A.43}$$

Now we construct a simulator that is a combination between a smooth simulator and a frequency simulator. The simulator $\tilde{l}(h_n|w_n)$ becomes

$$\frac{P(\nu \in \tilde{Z}^*)}{R} \sum_{r=1}^{R} I(\nu^*_{nr} \in Z^*) \chi(h_n|\nu^*_{nr}, w_n) k(\nu^*_{nr}, \gamma_1, \gamma_2, \rho) \tag{A.44}$$

Until now, we have considered the integration of the integrals appearing in the numerator and denominator of the score contribution separately. However, it is possible to construct a simulator on the basis of the vector of scores which has expectation zero in the true parameter vector. The drawback of this simulator is that we have to draw the random preference variables from their conditional density, i.e. conditional on the observed value of labour supply. The density function conditional on labour supply contains the same integral which we want to avoid to evaluate. Hence, we will consider an approximation. First consider the score of the log-likelihood contribution with respect to parameters that appear in the integrand only. Note that the derivatives of $g(v, \gamma_1, \gamma_2)$ with respect to its parameters can always be written as a multiple of $g(v, \gamma_1, \gamma_2)$ and therefore they can be treated in the same way as the following case.

The derivative of $l(h_n|w_n)$ with respect to σ^2_ϵ is

$$\frac{\partial l(h_n|w_n)}{\partial \sigma^2_\epsilon} = \int_{Z^*} \frac{\partial \chi(h_n|v, w_n)}{\partial \sigma^2_\epsilon} g(v, \gamma_1, \gamma_2) dv \tag{A.45}$$

and

$$\frac{\partial \ln l(h_n|w_n)}{\partial \sigma^2_\epsilon} = \frac{\partial l(h_n|w_n)}{\partial \sigma^2_\epsilon} / l(h_n|w_n) \tag{A.46}$$

For expository purposes, for the moment we ignore again the problem of the unknown bounds of Z^*. Now suppose that v^*_{nr} can be drawn from

$$\frac{g(v, \gamma_1, \gamma_2) \chi(h_n|v, w_n)}{l(h_n|w_n)}, v \in Z^* \tag{A.47}$$

Then the score contribution can be simulated by

$$\frac{\widetilde{\partial \ln} l(h_n|w_n)}{\partial \sigma^2_\epsilon} = P(v \in Z^*) \frac{1}{R} \sum_{r=1}^{R} \frac{\partial \chi(h_n|v^*_{nr}, w_n)}{\partial \sigma^2_\epsilon} / \chi(h_n|v^*_{nr}, w_n) \tag{A.48}$$

Taking expectations with respect to the draws v^*_{nr}, this yields the original score component.

For parameters which appear in the bounds of the integral we also have to differentiate the bounds. Suppose that the bound of the region Z^* is given by b. Optimal labour supply, evaluated in the bound is zero. Taking the derivatives of the bound yields

$$\pm \frac{\partial b}{\partial \vartheta}\phi(h_n, \sigma_\epsilon^2) g(b, \gamma_1, \gamma_1) / l(h_n|w_n) \tag{A.49}$$

where the sign is positive or negative, depending on whether it is an upperbound or a lowerbound. Taking expectations with respect to positive values of observed labour supply, i.e. with respect to the density $l(h_n|w_n)/P(h_n > 0)$ yields:

$$\pm \frac{1}{2}\frac{\partial b}{\partial \vartheta} g(b, \gamma_1, \gamma_1)/P(h_n > 0) = \frac{1}{2}\frac{\partial \int_{Z^*} g(v, \gamma_1, \gamma_2) dv}{\partial \vartheta}/P(h_n > 0) \tag{A.50}$$

which is one half times the derivative of the probability that optimal labour supply exceeds zero, divided by the probability that observed labour supply exceeds zero. Now the same trick can be applied as in Bloemen and Kapteyn (1993), which means that the original derivative of the bound is replaced by one half times the probability that optimal labour supply is greater than zero for every individual in the sample, both working and non-working. The derivative of the probability that optimal labour supply is zero can be simulated in the same way as in (A.35). Summarizing, for parameters which appear in the bound as well as in the integrand we can use the same kind of simulator as in (A.48) and in addition to that, we have to adjust the score contribution with a simulator of the derivative of the probability that optimal labour supply is positive for both working and non-working individuals. The resulting score simulator has expectation zero in the true parameter vector. For the problem of the unobserved bounds of Z^*, the same method can be used as in (A.44).

The practical appliability of this score method is restricted by the fact that random draws from the conditional density are required. The method, however, still has the advantage that the discontinuities are averaged out as the simulated score consists of linear contributions. Therefore it is useful to draw random number from an approximate density. The most straightforward way to draw random numbers in this context is to draw them from their marginal distribution. In that case the score contribution will not be unbiased anymore, i.e. the appropriate moment conditions are used in combination with the wrong draws. This point can be made more clear if for the moment we ignore the tax system. Suppose that optimal labour supply is given by the function $h_n(v_n)$ for individual n. The participation scheme (2.16) and (2.17) for observed labour supply is

$$\begin{aligned} h_n &= h_n(v_n) + \epsilon \quad \text{if } h_n(v_n) > 0 \text{ and } h_n(v_n) + \epsilon > 0 \\ &= \quad 0 \qquad\qquad \text{if } h_n(v_n) = 0 \text{ or } h_n(v_n) + \epsilon \leq 0 \end{aligned} \tag{A.51}$$

and

$$\epsilon_n \sim \phi(\epsilon_n, \sigma_\epsilon^2) \tag{A.52}$$

as before. Then for positive h_n we have

$$h_n \sim \int_{Z^*} \phi(h_n - h_n(v), \sigma_\epsilon^2) g(v) dv / P(h_n > 0) \tag{A.53}$$

in which Z^* is, as before, the region for which optimal labour supply is positive. Now draw v^*_{nr} from its marginal density $g(.)$, whereas h_n can be considered as a draw from (A.53). Then the implied density for $\epsilon^*_{nr} := h_n - h_n(v^*_{nr})$ is

$$\int_{Z^*}\int_{Z^*} \phi(\epsilon^*_{nr} + h_n(\tilde{v}) - h_n(v), \sigma^2_\epsilon)g(v)g(\tilde{v})dvd\tilde{v}/[P(\tilde{v} \in Z^*)P(h_n > 0)] \tag{A.54}$$

which is not equal to $\phi(\epsilon^*_{nr}, \sigma^2_\epsilon)/P(h_n > 0)$. Rather, it is a weighted average of $\phi(\epsilon_n, \sigma^2_\epsilon)/P(h_n > 0)$. Note that the implicit draws ϵ^*_{nr} are always in the right region, i.e. $h_n(v^*_{nr}) + \epsilon^*_{nr} > 0$. It can be shown that the parameter estimate for the variance of measurement error is biased upwards.

The Monte Carlo results in chapter 6 are obtained with method (A.48) using draws form the marginal density of v. From the results it can be seen that the estimate of the variance of measurement error is indeed biased upwards. However, the estimates of the utility parameters do rather well.

References

Bloemen, H.G. and A. Kapteyn (1993), *The joint estimation of a non-linear labour supply function and a wage equation using simulated response probabilities*, Annales d'Economie et de Statistiques, Nr. 29, pp. 175-205.

Blomquist, N. (1983), *The effect of income taxation on the labour supply of married men in Sweden*, Journal of Public Economics, Vol. 22, pp. 169-197.

Blundell, R. and C. Meghir (1987), *Bivariate alternatives to the Tobit model*, Journal of Econometrics, Vol. 34, pp. 179-200.

Blundell, R.W, J. Ham and C. Meghir (1987), *Unemployment and Female Labour Supply*, Economic Journal, 97, pp. 44-64

Börsch-Supan, A. and V. Hajivassiliou (1993), *Smooth unbiased multivariate probability simulators for maximum likeihood estimation of limited dependent variable models*, Journal of Econometrics, Vol. 58, nr. 3, pp. 347-368.

Devroye, L. (1986), *Non-uniform random variate generation*, Springer Verlag, New York.

Gourieroux, C., J. Laffont and A. Montfort (1980),*Coherency conditions in simultaneous linear equation models with enodgenous switching regimes*, Econometric 48, pp. 675-695.

Hajivassiliou, V. (1991), *Simulation Estimation Methods for Limited Dependent Variable Models*, in: Handbook of Statistics, Vol. 11 (Econometrics), (forthcoming), North-Holland.

Hammersley, J.M. and D.C. Handscomb (1964), *Monte Carlo Methods*, London: Methuen.

Hausman, J. (1979), *The econometrics of labor supply on convex budget sets*, Economics letters 3, pp. 171-174.

Hausman, J. (1980), *The Effect of Wages, Taxes and Fixed Costs on Women's Labor Force Participation*, Journal of Public Economics 14 pp. 161-194

Hausman, J. (1981), *Labor supply*, in: How Taxes Affect Economic Behavior, ed. H. J. Aaron and J. A. Pechman, The Brookings Inst., Washington DC, pp. 27-83

Hausman, J. (1985), *The econometrics of nonlinear budget sets*, Econometrica 53, pp. 1255- 1282.

Hausman, J. and P. Ruud (1984), *Family labor supply with taxes*, American Economic Review, 74, pp. 242-248.

Kapteyn, A., P. Kooreman and A. van Soest (1990), *Quantity Rationing and Concavity in a Flexible Household Labor Supply Model*, The Review of Economics and Statistics, Vol. 72, pp. 55-62

Kloek, T. and H.K. van Dijk (1978), *Bayesian estimates of equation system parameters: an application of integration by Monte Carlo*, Econometrica, vol. 46, no. 1, pp. 1-19.

Kooreman, P. and A. Kapteyn (1986), *Estimation of rationed and unrationed household labor supply functions using flexible functional forms*, The Economic Journal, Vol. 96, pp. 398-412.

Lerman, C. and C. Manski (1981), *On the Use of Simulated Frequencies to Approximate Choice Probabilities*, in: Structural Analysis of Discrete Data with Economic Applications, ed. C. Manski and D. McFadden, MIT Press, Cambridge, pp. 305-319

MaCurdy, T, D. Green and H. Paarsch (1990), *Assessing empirical approaches for analyzing taxes and labor supply*, Journal of Human Resources, vol. 25, nr. 3, pp. 415-490.

McFadden, D. (1989), *A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration*, Econometrica, Vol. 57, pp. 995-1026

Moffit, R. (1986), *The econometrics of piecewise-linear budget constraints*, Journal of Business and economic Statistics 4, pp. 317-328.

Nelder, J.A. and R. Mead (1965), *A simplex method for function minimization*, Computer Journal Vol. 7, pp. 308-313.

Pakes, A. and D. Pollard (1989), *Simulation and the Asymptotics of Optimization Estimators*, Econometrica, Vol. 57, pp. 1027-1057

Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling (1986), *Numerical Recipes*, Cambridge University Press, Cambridge

Van Soest, A., P. Kooreman and A. Kapteyn (1993), *Coherency and Regularity of Demand Systems with Equality and Inequality Constraints*, Journal of Econometrics, Vol. 57, nr. 1-3, pp. 161.

Discussion Paper Series, CentER, Tilburg University, The Netherlands:

(For previous papers please consult previous discussion papers.)

No.	Author(s)	Title
9232	F. Vella and M. Verbeek	Estimating the Impact of Endogenous Union Choice on Wages Using Panel Data
9233	P. de Bijl and S. Goyal	Technological Change in Markets with Network Externalities
9234	J. Angrist and G. Imbens	Average Causal Response with Variable Treatment Intensity
9235	L. Meijdam, M. van de Ven and H. Verbon	Strategic Decision Making and the Dynamics of Government Debt
9236	H. Houba and A. de Zeeuw	Strategic Bargaining for the Control of a Dynamic System in State-Space Form
9237	A. Cameron and P. Trivedi	Tests of Independence in Parametric Models: With Applications and Illustrations
9238	J.-S. Pischke	Individual Income, Incomplete Information, and Aggregate Consumption
9239	H. Bloemen	A Model of Labour Supply with Job Offer Restrictions
9240	F. Drost and Th. Nijman	Temporal Aggregation of GARCH Processes
9241	R. Gilles, P. Ruys and J. Shou	Coalition Formation in Large Network Economies
9242	P. Kort	The Effects of Marketable Pollution Permits on the Firm's Optimal Investment Policies
9243	A.L. Bovenberg and F. van der Ploeg	Environmental Policy, Public Finance and the Labour Market in a Second-Best World
9244	W.G. Gale and J.K. Scholz	IRAs and Household Saving
9245	A. Bera and P. Ng	Robust Tests for Heteroskedasticity and Autocorrelation Using Score Function
9246	R.T. Baillie, C.F. Chung and M.A. Tieslau	The Long Memory and Variability of Inflation: A Reappraisal of the Friedman Hypothesis
9247	M.A. Tieslau, P. Schmidt and R.T. Baillie	A Generalized Method of Moments Estimator for Long-Memory Processes

No.	Author(s)	Title
9248	K. Wärneryd	Partisanship as Information
9249	H. Huizinga	The Welfare Effects of Individual Retirement Accounts
9250	H.G. Bloemen	Job Search Theory, Labour Supply and Unemployment Duration
9251	S. Eijffinger and E. Schaling	Central Bank Independence: Searching for the Philosophers' Stone
9252	A.L. Bovenberg and R.A. de Mooij	Environmental Taxation and Labor-Market Distortions
9253	A. Lusardi	Permanent Income, Current Income and Consumption: Evidence from Panel Data
9254	R. Beetsma	Imperfect Credibility of the Band and Risk Premia in the European Monetary System
9301	N. Kahana and S. Nitzan	Credibility and Duration of Political Contests and the Extent of Rent Dissipation
9302	W. Güth and S. Nitzan	Are Moral Objections to Free Riding Evolutionarily Stable?
9303	D. Karotkin and S. Nitzan	Some Peculiarities of Group Decision Making in Teams
9304	A. Lusardi	Euler Equations in Micro Data: Merging Data from Two Samples
9305	W. Güth	A Simple Justification of Quantity Competition and the Cournot-Oligopoly Solution
9306	B. Peleg and S. Tijs	The Consistency Principle For Games in Strategic Form
9307	G. Imbens and A. Lancaster	Case Control Studies with Contaminated Controls
9308	T. Ellingsen and K. Wärneryd	Foreign Direct Investment and the Political Economy of Protection
9309	H. Bester	Price Commitment in Search Markets
9310	T. Callan and A. van Soest	Female Labour Supply in Farm Households: Farm and Off-Farm Participation
9311	M. Pradhan and A. van Soest	Formal and Informal Sector Employment in Urban Areas of Bolivia
9312	Th. Nijman and E. Sentana	Marginalization and Contemporaneous Aggregation in Multivariate GARCH Processes
9313	K. Wärneryd	Communication, Complexity, and Evolutionary Stability

No.	Author(s)	Title
9314	O.P.Attanasio and M. Browning	Consumption over the Life Cycle and over the Business Cycle
9315	F. C. Drost and B. J. M. Werker	A Note on Robinson's Test of Independence
9316	H. Hamers, P. Borm and S. Tijs	On Games Corresponding to Sequencing Situations with Ready Times
9317	W. Güth	On Ultimatum Bargaining Experiments - A Personal Review
9318	M.J.G. van Eijs	On the Determination of the Control Parameters of the Optimal Can-order Policy
9319	S. Hurkens	Multi-sided Pre-play Communication by Burning Money
9320	J.J.G. Lemmen and S.C.W. Eijffinger	The Quantity Approach to Financial Integration: The Feldstein-Horioka Criterion Revisited
9321	A.L. Bovenberg and S. Smulders	Environmental Quality and Pollution-saving Technological Change in a Two-sector Endogenous Growth Model
9322	K.-E. Wärneryd	The Will to Save Money: an Essay on Economic Psychology
9323	D. Talman, Y. Yamamoto and Z. Yang	The $(2^{n+m+1} - 2)$-Ray Algorithm: A New Variable Dimension Simplicial Algorithm For Computing Economic Equilibria on $S^n \times R^m_+$
9324	H. Huizinga	The Financing and Taxation of U.S. Direct Investment Abroad
9325	S.C.W. Eijffinger and E. Schaling	Central Bank Independence: Theory and Evidence
9326	T.C. To	Infant Industry Protection with Learning-by-Doing
9327	J.P.J.F. Scheepens	Bankruptcy Litigation and Optimal Debt Contracts
9328	T.C. To	Tariffs, Rent Extraction and Manipulation of Competition
9329	F. de Jong, T. Nijman and A. Röell	A Comparison of the Cost of Trading French Shares on the Paris Bourse and on SEAQ International
9330	H. Huizinga	The Welfare Effects of Individual Retirement Accounts
9331	H. Huizinga	Time Preference and International Tax Competition
9332	V. Feltkamp, A. Koster, A. van den Nouweland, P. Borm and S. Tijs	Linear Production with Transport of Products, Resources and Technology

No.	Author(s)	Title
9333	B. Lauterbach and U. Ben-Zion	Panic Behavior and the Performance of Circuit Breakers: Empirical Evidence
9334	B. Melenberg and A. van Soest	Semi-parametric Estimation of the Sample Selection Model
9335	A.L. Bovenberg and F. van der Ploeg	Green Policies and Public Finance in a Small Open Economy
9336	E. Schaling	On the Economic Independence of the Central Bank and the Persistence of Inflation
9337	G.-J. Otten	Characterizations of a Game Theoretical Cost Allocation Method
9338	M. Gradstein	Provision of Public Goods With Incomplete Information: Decentralization vs. Central Planning
9339	W. Güth and H. Kliemt	Competition or Co-operation
9340	T.C. To	Export Subsidies and Oligopoly with Switching Costs
9341	A. Demirgüç-Kunt and H. Huizinga	Barriers to Portfolio Investments in Emerging Stock Markets
9342	G.J. Almekinders	Theories on the Scope for Foreign Exchange Market Intervention
9343	E.R. van Dam and W.H. Haemers	Eigenvalues and the Diameter of Graphs
9344	H. Carlsson and S. Dasgupta	Noise-Proof Equilibria in Signaling Games
9345	F. van der Ploeg and A.L. Bovenberg	Environmental Policy, Public Goods and the Marginal Cost of Public Funds
9346	J.P.C. Blanc and R.D. van der Mei	The Power-series Algorithm Applied to Polling Systems with a Dormant Server
9347	J.P.C. Blanc	Performance Analysis and Optimization with the Power-series Algorithm
9348	R.M.W.J. Beetsma and F. van der Ploeg	Intramarginal Interventions, Bands and the Pattern of EMS Exchange Rate Distributions
9349	A. Simonovits	Intercohort Heterogeneity and Optimal Social Insurance Systems
9350	R.C. Douven and J.C. Engwerda	Is There Room for Convergence in the E.C.?
9351	F. Vella and M. Verbeek	Estimating and Interpreting Models with Endogenous Treatment Effects: The Relationship Between Competing Estimators of the Union Impact on Wages

No.	Author(s)	Title
9352	C. Meghir and G. Weber	Intertemporal Non-separability or Borrowing Restrictions? A Disaggregate Analysis Using the US CEX Panel
9353	V. Feltkamp	Alternative Axiomatic Characterizations of the Shapley and Banzhaf Values
9354	R.J. de Groof and M.A. van Tuijl	Aspects of Goods Market Integration. A Two-Country-Two -Sector Analysis
9355	Z. Yang	A Simplicial Algorithm for Computing Robust Stationary Points of a Continuous Function on the Unit Simplex
9356	E. van Damme and S. Hurkens	Commitment Robust Equilibria and Endogenous Timing
9357	W. Güth and B. Peleg	On Ring Formation In Auctions
9358	V. Bhaskar	Neutral Stability In Asymmetric Evolutionary Games
9359	F. Vella and M. Verbeek	Estimating and Testing Simultaneous Equation Panel Data Models with Censored Endogenous Variables
9360	W.B. van den Hout and J.P.C. Blanc	The Power-Series Algorithm Extended to the *BMAP/PH/*1 Queue
9361	R. Heuts and J. de Klein	An (s,q) Inventory Model with Stochastic and Interrelated Lead Times
9362	K.-E. Wärneryd	A Closer Look at Economic Psychology
9363	P.J.-J. Herings	On the Connectedness of the Set of Constrained Equilibria
9364	P.J.-J. Herings	A Note on "Macroeconomic Policy in a Two-Party System as a Repeated Game"
9365	F. van der Ploeg and A. L. Bovenberg	Direct Crowding Out, Optimal Taxation and Pollution Abatement
9366	M. Pradhan	Sector Participation in Labour Supply Models: Preferences or Rationing?
9367	H.G. Bloemen and A. Kapteyn	The Estimation of Utility Consistent Labor Supply Models by Means of Simulated Scores

P.O. BOX 90153, 5000 LE TILBURG, THE NETHERLAND