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

by

Hans G. Bloemen* and Arie Kapteyn**

Address for correspondence:
Free University Amsterdam
Department of Economics
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands

This version September 2003

Abstract

We consider 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. Coherence 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 experiments and for the real data the various estimation methods yield very different results. Furthermore estimates are sensitive to the exact specification of the budget constraint.

---

1 Strategisch Arbeidsmarktonderzoek (OSA) for kindly providing the data. Thanks are due to Arthur van Soest and two anonymous referees for their comments. Email addresses: * hbloemen@feweb.vu.nl (Free University Amsterdam), ** Arie_Kapteyn@rand.org (Rand Corporation)
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 Maximum Likelihood (ML) to the Method of Moments (MM), all estimation methods applied in practice can be seen as approximations with varying degrees of accuracy.

Rather closely related to the previous issues is the issue of coherency.\(^3\) It turns out that in models with kinked budget constraints coherency requires quasiconcavity of the direct utility function at all kink points.\(^4\) Since these kink points will be different for different

---

\(^2\) 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.

\(^3\) 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.

\(^4\) More specifically, in the context of labour supply models with kinked budget constraints the consequence of incoherence is that for given parameter values the budget constraint may be tangent to the indifference curve at more than one (kink) point, thus generating several values for an agent’s optimal working hours. A statistical implication of incoherence is that the probabilities of working and non-working add up to more than 1. The latter causes parameter estimates to be inconsistent if coherency is not imposed (see Kooreman, Van Soest and Kaptchyn (1993)). Note that coherency is a different concept than the identification of model parameters: model parameters are not identified if different parameter values yield the same optimal working hours or participation probability.
individuals in a sample, coherency requires quasi-concavity at many combinations of hours and wages. See MaCurdy, Green, Paarsch (1990) or Van Soest, Kooiman, 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.\textsuperscript{5} Except for simple models, the imposition of utility consistency is non-trivial. Failure to do so, however, may lead to inconsistent estimators.

In this paper we attack the estimation of a utility consistent static model of individual labor supply, with a flexible specification of preferences head on. 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), originally developed by Hajivassiliou and McFadden (1998). By implementing this estimation method we incorporate all the features of the model model specified: we account for the stochastic specification, for the kinked budget constraint and for the coherency constraint. By using simulators for the participation probabilities 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 (1993). 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.\textsuperscript{6} Although specific assumptions will be made regarding functional forms and stochastic specifications, the approach is general, and can be applied under a variety of alternative assumptions.

\textsuperscript{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.

\textsuperscript{6} It should be noted that in specific applications frequency simulators can be replaced by smoothed variants of the frequency simulator. For instance, Keane and Moffit (1998) use specific properties of their specification to employ a smoothed simulator. To fully test for the the generality of simulation methods in this context, we will employ a frequency simulator throughout.
Since the application of the MSS method is quite involved, it is tempting to use alternative methods of estimation that are much easier to implement. Possible alternatives include the estimation by IV using a two-step procedure, the use of predicted wages, and ignoring randomness of preferences (or equivalently, unobserved heterogeneity). In the context of a utility consistent labour supply model with a flexible specification of preferences and a nonlinear budget constraint, these methods are generally not consistent with the model assumptions. However, if these approximate methods would yield estimates that hardly differ from MSS, there would be no objection against their implementation. For this reason we compare the outcome of the MSS method with various approximate methods.

The order of presentation is as follows: In the next section we set out our model: we present the specification, the restrictions that have to be imposed to render the model coherent and we show how simulation methods can be used in the particular context of the model. In section 3 we present the estimation by the method of simulated scores. In section 4 we discuss various approaches in the literature using the model as an illustration and subsequently we compare our estimation strategy 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 5 concludes.

2 Model specification

2.1 Specification of preferences

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

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

(2.1)

where

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

(2.2)

The variables \( h \) and \( c \) are hours worked and consumption respectively; \( \beta, \gamma, \delta, \theta \) are parameters. Maximization of (2.1) subject to a linear budget constraint of the form \( c \leq w h + \mu \), where \( w \) is the wage rate and \( \mu \) is nonlabor income, yields the following labor supply function:

\[ h(w, \mu) = \delta + \mu^* \beta + w \gamma \]  

(2.3)

with

\[ \mu^* = \theta + \mu + \delta w + \frac{1}{2} \gamma w^2 \]  

(2.4)
We will assume throughout that $\gamma > 0$. It is easy to see that convexity of the indifference curves requires
\[
m^* < \frac{\gamma}{\beta^2}
\]
(2.5)

Notice that under this condition utility is increasing in consumption.

Using (2.2), (2.5) can be written as follows:
\[
\left\{ 1 + \frac{\beta^2}{\gamma} \left[ \frac{(h - \delta)^2}{\gamma} - 2(c + \theta) \right] \right\}^{\frac{1}{2}} > 0
\]
(2.6)

Clearly this holds true whenever the utility function is defined. The condition for the existence of the utility function can be written as:
\[
c < -\theta + \frac{\gamma}{2\beta^2} + \frac{(h - \delta)^2}{2\gamma} \equiv f(h)
\]
(2.7)

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 $U(h, c)$.

2.2 Stochastic specification

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:

\[
\ln w_n = \eta' x_n + u_n \tag{2.8}
\]

where \( u_n \) is an error term representing unobserved heterogeneity, \( x_n \) is a vector of observable characteristics and \( \eta \) is a vector of parameters. As to the labor supply equation, we introduce preference variation by allowing \( \theta \) to vary across agents as follows:

\[
\theta_n = \theta_0 + q_n^\prime \omega + v_n \tag{2.9}
\]

where \( v_n \) is an unobservable error term; \( \omega \) is a parameter vector; and \( q_n \) is a vector of observable characteristics, which, due to exclusion restrictions, is a subset of \( x_n \) in (2.8).\(^7\)

For later purposes it is useful to define also

\[
\mu^*_n = \theta_n + \mu_n + w_n \delta + \frac{1}{2} w_n^2 \gamma
\]

For given values of \( w_n \) and \( v_n \) individual \( n \)'s optimal number of hours, say \( \overline{h}_n \), is determined by the maximization of utility subject to a non-linear and non-convex budget set. Figure 3 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 \textit{maximum maximorum}.

In general this results in a complicated function of the wage, nonlabor income, individual characteristics and the random preference term. We write this labour supply function as

\[
\overline{h}_n = h_n(w_n, \mu_n; \alpha, q_n^\prime \omega + v_n) \tag{2.11}
\]

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

\[
h^*_n = \overline{h}_n + \epsilon_n \tag{2.12}
\]

Let \( h_n \) be observed labor supply, then we assume:

\[
\begin{align*}
h_n &= h^*_n \quad \text{if } h^*_n > 0 \quad \text{and } h_n(w_n, \mu_n; \alpha, q_n^\prime \omega + v_n) > 0 \tag{2.13} \\
h_n &= 0 \quad \text{if } h^*_n \leq 0 \quad \text{or } h_n(w_n, \mu_n; \alpha, q_n^\prime \omega + v_n) \leq 0 \tag{2.14}
\end{align*}
\]

\(^7\) There is no a priori reason to let preference variation enter through \( \theta \) 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.
This formulation brings out the distinction between the random preferences \( v_n \), which, among other things, influence the probability that \( \pi_n \) is positive and the optimization or measurement errors \( \epsilon_n \), which, given \( \pi_n \), influence the probability of observing positive labor supply. To appreciate the specification (2.13)-(2.14), it is useful to interpret \( \epsilon_n \) as a deviation from optimal hours due to demand side restrictions. These restrictions may arise because no job is available with exactly the optimal number of hours, or because no job is available at all. Under this interpretation (2.13) states that if both the optimal number of hours and the feasible number of hours are positive, we will observe the individual to be working the feasible number of hours. Equation (2.14) states that we will observe an individual not to be working if the optimal number of hours is negative or if no job with a positive number of hours is available. This case therefore covers both non-participation and involuntary unemployment.\(^8\) The assumption then is that involuntary employment does not occur (i.e., optimal hours are negative, and still the individual would work). If \( \epsilon_n \) represents pure measurement error, (2.13) and (2.14) imply an asymmetry: someone who is actually working, may be recorded as not working in our data, but someone who is not working in reality will not be observed as working in our data (see below).

The introduction of the stochastic terms affects the condition for the convexity of the utility function. In particular, we want (2.7) to hold true for all datapoints. To indicate this, we add subscripts and write:

\[
\epsilon_n < -\theta_n + \frac{\gamma}{2\beta^2} + \frac{(h_n - \delta)^2}{2\gamma} \equiv f_n(h_n) \tag{2.15}
\]

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. 2 for an illustration.

If we restrict the range of \( \theta_n \) such that inequality (2.15) holds for all values of \( \epsilon_n \) 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 \( \theta_n \) 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

\[
\theta_n = -\tilde{\mu} - \delta \tilde{w} - \frac{1}{2} \gamma \tilde{w}^2 + \frac{\gamma}{2\beta^2} \tag{2.16}
\]

\(^8\) In the empirical work below we will also consider more general formulations to allow for involuntary unemployment. For the present discussion such a generalization is not necessary.
Thus, in view of (2.15) the inequality constraint on $\theta_n$ has to be

$$\theta_n < -\bar{\mu} - \delta \hat{w} - \frac{1}{2} \gamma \hat{w}^2 + \frac{\gamma}{2 \beta^2}$$

(2.17)

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

$$\theta_0 < -q'_n \omega - \bar{\mu} - \delta \hat{w} - \frac{1}{2} \gamma \hat{w}^2 + \frac{\gamma}{2 \beta^2}$$

(2.18)

for all $n$. For the random preference term $v_n$ we will actually assume that it follows a negative $\Gamma$ 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, which we define over the gross wage rate, to $[0, \bar{w}]$. The upperbound $\bar{w}$ on the gross wage rate is a transformation of $\hat{w}$ that involves the appropriate marginal tax rate. A convenient choice of distribution for $u_n$, which restricts the range of $w_n$ to $[0, \bar{w}]$, is to define a random variable $\lambda_n$ following a lognormal distribution with log-mean $m_n$ and log-variance $\tau^2$, and to define

$$u_n \equiv \log\{\psi_n \lambda_n/[1 + \lambda_n]\}$$

(2.19)
where
\[ \psi_n \equiv \bar{w} / \exp[w(x_n, \eta)] \]  

(2.20)

It seems reasonable to require that the median of \( \exp(u_n) \) equals one. This holds true if we specify
\[ m_n = - \log(\psi_n - 1) \]

(2.21)

Thus the number of free parameters is equal to the case where the \( u_n \) were assumed normal with mean zero.\(^9\)

### 2.3 Simulation of working hours and probabilities

We have already noted that the labour supply function (2.11) is defined implicitly and need not be continuous. As a consequence, the density function of observed labour supply will be difficult to evaluate as well. However, it is fairly straightforward to simulate labour supply and the participation probabilities, as we will make clear now. Most of the simulation can be understood by reconsidering Fig. 3.

\(^9\) Notice that under normality the inequality (2.18) is violated with non-zero probability and hence the model would be incoherent.
Figure 3 presents an example of a non-linear budget constraint and a non-convexity. 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 euro of labour income, say, \( \alpha \% \) of the unemployment benefit is taxed away. At hours \( H_0 \), the individual receives no benefit anymore. The virtual wage rate on the first segment is indicated by \( w_0 \). On the second segment, between \( H_0 \) and \( H_1 \), no more benefits are received and the virtual wage rate rises to \( w_1 \). At hours \( H_1 \) the next tax bracket is reached, which is reflected by a fall in the virtual wage rate to \( w_2 \) on the third segment. Now let \( h_{ij}^n \) denote the optimal labour supply of individual \( n \) at a linear budget constraint with slope \( w_{nj} \) and intercept \( \mu_{nj} \), \( j = 0, 1, \ldots, m \), and denote optimal labour supply by \( \tilde{h}_n \), which is generated by the labour supply function (2.11). Note that in (2.11) \( w_n \) is the before tax wage rate which implies that the function \( h(.) \) includes the tax and the welfare system. Then the optimal labour supply \( \tilde{h}_n \), conditional on \( v_n \) and \( w_n \), can be determined according to the following algorithm:

\[
\begin{align*}
    h_{n,NC} &= 0 \quad \text{if} \quad \begin{array}{c}
        h_0^n \\
        h_0^n \\
        H_0 \\
        H_0 \\
        h_j^n \\
        H_{n,j-1} \\
        H_{n,j} \\
        H_{n,m-1} \\
    \end{array} \quad \begin{array}{c}
        \leq 0 \\
        \leq H_0 \\
        > H_0 \\
        < H_0 \\
        < h_{j+1}^n \\
        \leq H_{n,j} \\
        > h_{n,j} \\
        \leq T \\
    \end{array} \\
    \tilde{h}_n &= \begin{array}{c}
        \tilde{h}_n \\
        \in [0, H_0] \\
        \in (H_0, H_{n,j}) \\
        \in (H_{n,j}, h_{n,j}^m) \\
        \in (h_{n,j}^m, T) \\
    \end{array} \\
    \tilde{h}_n &= \begin{array}{c}
        \text{if utility in } \tilde{h}_n \text{ exceeds utility in } h_{n,NC} \\
        \text{otherwise} \\
    \end{array} \\
    h_j^n &= \alpha_1 + \alpha_2 \mu_{nj} + \alpha_3 w_{nj} + \frac{1}{2} \alpha_4 w_{nj}^2 + q_n' \zeta + \alpha_2 v_n
\end{align*}
\]

The parameters \( \alpha \) and \( \zeta \) are obtained by reparameterization of \( \beta, \gamma, \delta \) and \( \omega \) in section 2. The precise form of this reparameterization is given in the appendix. \( h_{n,NC} \) is the number of hours with maximum utility on the segment \((0, H_0)\), i.e., the segment that causes the budget set to be non-convex (NC). The number of hours yielding maximum utility on the segment \((H_0, T)\) is denoted by \( \tilde{h}_n \). Optimal labour supply \( \tilde{h}_n \) is determined by comparing the utility values of these two segments and choosing the maximum maximorum. Note that the latter choice involves the calculation of the direct utility function on the two segments of the budget set, which is something that is not necessary in the case of a piecewise linear and convex budget constraint.
Variation in the gross wage rate $w_n$ and in the unobservable taste component $v_n$, determine the segments of the budget constraint: $w_n$ determines the slopes and the kink points, and $v_n$ determines in which segment labour supply will be optimal, given everything else. Now the procedure for constructing the density of observed working hours $h_n$ would be as follows. First, the distribution of measurement or optimization errors $\epsilon_n$ can be used to determine the distribution of $h_n^*$, defined in (2.12), conditional on the unobserved taste component $v_n$ and on the gross wage rate $w_n$. Next, the distribution of $v_n$ can be used to integrate out the unobserved taste component, taking into account the decision rule on $\bar{h}_n$ in (2.13) and (2.14). Note, though, that the region of integration is in general only defined implicitly, since integration involves comparing utility values and integrating over utility functions. Finally, multiplication with the marginal density of the gross wage rate is required to determine the joint density of hours and wages.

Evaluating the probability of nonworking involves the integration over wage rates, in addition to the integration over random preferences. It is clear that the expression for the probability that observed labour supply equals zero ($h_n = 0$) is complicated and it is difficult even to write down an analytic expression for the probability and the likelihood function as a whole (see Appendix A). 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 of zero observed labor supply with a frequency simulator $F_{nR}$, which is constructed as follows: Draw $R$ times a wage rate $w_{nr}^*$, an unobserved taste variable $v_{nr}^*$ and a measurement error $\epsilon_{nr}^*$ from their assumed distributions, calculate $\bar{h}_n$ and $h_n^*$ and use the rules in (2.13) and (2.14) to determine $h_n$. The simulator becomes:

$$f_{nr} = 1 \quad \text{if } h_n \leq 0$$  
$$f_{nr} = 0 \quad \text{otherwise}$$  

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

In the next section we will describe how the frequency simulator of the probability can be implemented in a method of simulated scores estimation procedure.
3 Estimation

3.1 The method

For ease of notation we introduce the dummy variable $d_n$ with

\[
\begin{align*}
    d_n &= 1 \quad \text{if} \quad h_n = 0 \\
    d_n &= 0 \quad \text{if} \quad h_n > 0
\end{align*}
\] (3.1)

Furthermore we write $P_n(\vartheta)$, where $\vartheta$ contains the parameters of $\alpha$, $\eta$ and the parameters of the distribution functions of $u_n$, $v_n$, and $\epsilon_n$, for the probability that $h_n$ equals zero.\(^{10}\)

The joint density of hours $h_n$ and wages $w_n$ is a mixed discrete-continuous probability density function, which we denote by $g(h_n, w_n|x_n, \mu_n, \vartheta)$:

\[
g(h_n, w_n|x_n, \mu_n, \vartheta) = \begin{cases} 
    P_n(\vartheta) & \text{if } h_n = 0 \\
    g^*(h_n, w_n|x_n, \mu_n, \vartheta) & \text{if } h_n > 0, 0 < w_n < \infty
\end{cases}
\] (3.2)

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 $\bar{P}_n$.

Generally, the log-likelihood function of the model is

\[
L(\vartheta|x_n, \mu_n, w_n, h_n, n = 1, \ldots, N) = \\
\sum_{n=1}^{N} [d_n \ln P_n(\vartheta) + (1 - d_n) \ln g^*(h_n, w_n|x_n, \mu_n, \vartheta)] 
\] (3.3)

It will be assumed that the likelihood is differentiable almost everywhere with respect to the elements of $\vartheta$.\(^{11}\) Thus we can differentiate the log-likelihood with respect to $\vartheta$ to derive the first order conditions for a maximum.

\[
\frac{\partial L(\vartheta)}{\partial \vartheta} = \sum_{n=1}^{N} \left[ d_n \frac{\partial \ln P_n(\vartheta)}{\partial \vartheta} + (1 - d_n) \frac{\partial \ln g^*(h_n, w_n|x_n, \mu_n, \vartheta)}{\partial \vartheta} \right] = \sum_{n=1}^{N} \frac{\partial L_n(\vartheta)}{\partial \vartheta} 
\] (3.4)

The maximum likelihood estimator $\hat{\vartheta}_{ML}$ of $\vartheta$ satisfies

\[
\frac{1}{N} \sum_{n=1}^{N} \frac{\partial L_n(\hat{\vartheta}_{ML})}{\partial \vartheta} = 0 
\] (3.5)

---

\(^{10}\) In the literature, the state of employment is often indicated with a dummy variable which takes the value one. Note that we do the reverse in our notation and indicate the state of nonworking with the value one. The reason for this is that in the estimation the probability of nonworking needs to be evaluated, so the expectation of our labour market state indicator $d_n$ naturally equals this probability, which turns out to be convenient in the further description of the properties of our estimator.

\(^{11}\) For consistency some additional regularity conditions are required: $\frac{\partial^2 L}{\partial \vartheta^2}$ exists in a neighbourhood of $\vartheta_0$ and is non-singular and negative definite in a neighbourhood of $\vartheta_0$. Moreover, $\vartheta$ is required to lie in a compact set for MLE to be consistent.
whereas for the true parameter value, denoted by \( \hat{\theta} \), it holds that the score vector has expectation zero:

\[
E \left( \frac{\partial L(\hat{\theta})}{\partial \hat{\theta}} \right) = 0.
\] (3.6)

This leads to the interpretation of the ML-estimator as an MM-estimator. Consistency of this MM-estimator follows from applying a law of large number argument and some further regularity conditions to (3.5) and (3.6).

In the present context the evaluation of the score vector is impossible for the reasons set out in the previous section. Following Hajivassiliou and McFadden (1998) we replace the score in (3.4) by an unbiased simulator \( K_R(\hat{\theta}) \equiv \sum_{n=1}^{N} K_n(\hat{\theta}) \). The subscript \( R \) indicates the number of replications that is used in the simulation. The unbiasedness of this simulated score implies that \( E K_R(\hat{\theta}) = E \frac{\partial L(\hat{\theta})}{\partial \hat{\theta}} \), so \( E K_R(\hat{\theta}) = 0 \). As a consequence, under certain regularity conditions (see below), this moment equation can be used for consistent estimation of \( \hat{\theta} \). Intuitively, this can be seen as follows. We can construct the sample moment

\[
\frac{1}{N} \sum_{n=1}^{N} K_n(\hat{\theta}) = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial L_n(\hat{\theta})}{\partial \hat{\theta}} + \frac{1}{N} \sum_{n=1}^{N} e_{nR},
\] (3.7)

where \( e_{nR} \) is the simulation error for the \( n \)-th observation, i.e., the difference between the score contribution and its simulator. By construction the \( e_{nR} \) are independent across observations. Hence, under mild regularity conditions \( 1/N \sum_{n=1}^{N} e_{nR} \) will converge in probability to zero for \( N \) going to infinity if the simulator of the score is unbiased.\(^{13} \)

This implies that the value of \( \hat{\theta} \) that sets the simulated score equal to zero will converge to the same value as the ML-estimator.

Since we want to use frequency simulators for the response probabilities, it is important that the probability enters the moment condition linearly, because this will allow the unbiased simulation of the probabilities within the moment condition. For this purpose we first rewrite the score vector (3.4) and next we show how it can be modified.

We may rewrite (3.4) as

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

where

\[
Z_n = \frac{\frac{\partial P_n}{\partial \hat{\theta}}}{P_n(1 - P_n)}.
\] (3.9)

\(^{12} \) A necessary assumption is that the support of \( h_n \) and \( w_n \) does not depend on \( \hat{\theta} \).

\(^{13} \) Note that for an unbiased simulator of the score the number of replications \( R \) does not influence the consistency of the moment estimator. In the case of simulated maximum likelihood, the (implicitly) simulated score is not unbiased. To obtain a \( \sqrt{N} \) consistent estimator usually requires that \( R/\sqrt{N} \to \infty \).
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 arbitrary vector of instruments $\tilde{Z}_n$, independent of $\vartheta$, the expectation of the resulting expression, conditional on $\tilde{Z}_n$, equals zero at the true parameter value $\vartheta_0$. Actually, this way of modifying the score is a direct extension of the method of simulated scores for discrete choice variables by McFadden (1989). In Section 2 we described how we simulate $P_n(\vartheta)$ unbiasedly with a frequency simulator.

Next, we use the fact that

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

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

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

The derivatives of $P_n(\vartheta)$ can be simulated using a difference approximation of frequency simulators which is an unbiased simulator for the difference approximation of the probabilities. We denote this simulator by $\tilde{m}_{nR}$.

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

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

where $\nu_{nR}^*$ is the vector of drawings used to simulate the score contribution for observation $n$.

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

$$K_R(\vartheta) = 0$$

A convenient way to evaluate the consistency of the estimator, is to look at the simulation errors, formed by the difference between the moment condition (3.11) and its simulated variant (3.12). The sample moment of the simulation residuals can be written as:

$$\frac{1}{NR} \sum_{n=1}^{N} \sum_{r=1}^{R} \left[ \tilde{Z}_n(P_n - k_{nr}) - \left\{ \tilde{m}_{nr} - (1 - d_n) \frac{\partial \ln \tilde{P}_n}{\partial \vartheta} \right\} \right]$$
with \( \frac{1}{R} \sum_{r=1}^{R} k_{nr} = k_n(\theta, v_{nR}^*) \) and \( \frac{1}{R} \sum_{r=1}^{R} \tilde{m}_{nr} = \tilde{m}_n(\theta, v_{nR}^*) \). Consistency of the estimator requires the simulation error to converge to zero as \( N \) tends to infinity. In case of continuous simulators standard asymptotic theory can be applied to prove consistency and to show that (3.14) converges to zero for fixed \( R \). There are two additional complications. The first is that we use frequency simulators so we do not have continuity. Pakes and Pollard (1989) used empirical process methods to prove the consistency of method of moments estimators is case of discontinuities. We comment on this in section 3.2. Up till now we have concentrated on the simulation of the probabilities. In our empirical application we will also simulate the derivative of the continuous part of the score by drawing random preferences: we simulate \( g^*(h_n, w_n|x_n, \mu_n, \theta) \) using the density function conditional on random preferences, say \( g^*_n(h_n, w_n|x_n, \mu_n, \theta) \). The computational details will be described in the appendix. Simulation of this density does not involve any problems of discontinuity, but it does introduce, like in simulated maximum likelihood, a denominator bias in the score vector. The simulation error is augmented accordingly. For \( \sqrt{N} \) consistency this of the estimator the additional condition is that \( R/\sqrt{N} \to \infty \).

### 3.2 Asymptotic properties of the estimator

The simulated score vector satisfies the property that its expectation, evaluated at the true parameter vector \( \theta_0 \), 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 \( \hat{\theta}_R \), at which the simulated score is zero, will converge to the true parameter value \( \theta_0 \), for \( N \to \infty \), or equivalently, \( \hat{\theta}_R \) will be a consistent estimator of \( \theta_0 \). 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 (3.12). Let \( G(\theta) \) denote the unconditional expectation of the term in square brackets in (3.12). \( G(\theta) \) is a vector of population moments. Note that \( G(\theta_0) = 0 \). Identifiability requires that \( \inf_{\|\theta-\theta_0\|>\delta} \| G(\theta) \| > 0 \ \forall \ \delta > 0 \). Furthermore, \( G(\theta) \) is required to have a non-singular derivative matrix at \( \theta_0 \).

\( \frac{1}{N} K_R(\theta) \) is the empirical counterpart of \( G(\theta) \). In their proof of consistency and asymptotic normality, Pakes and Pollard (1989) use the fact that \( \frac{1}{N} K_R(\theta) \) 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 \( G(\theta) \), making use of the independence-across-observations assumption. Consequently, no smoothness assumptions are required. A condition which
has to be satisfied when using a frequency simulator is that the probability of being at a tie (i.e., \( d = 1 \) and \( d = 0 \)) is zero at \( \vartheta_0 \). This condition is clearly satisfied here. Finally, Pakes and Pollard (1989) 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:

\[
\frac{1}{\sqrt{N}} K_R(\vartheta_0) \overset{\text{asy}}{\to} N(0, V_{R0}),
\]

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

\[
\sqrt{N}(\hat{\vartheta}_R - \vartheta_0) \overset{\text{asy}}{\to} N(0, \Gamma^{-1} V_{R0}(\Gamma')^{-1})
\]

where \( \Gamma' = \text{plim} \frac{1}{N} \frac{\partial L(\vartheta)}{\partial \vartheta} \) (3.17)

To estimate the covariance matrix we calculate

\[
\hat{\Omega}_R = \hat{\Gamma}^{-1} \hat{V}_R(\hat{\Gamma}')^{-1}
\]

with

\[
\hat{\Gamma}' = \frac{1}{N} \frac{\partial L(\vartheta)}{\partial \vartheta}
\]

(3.19)

\[
\hat{V}_R = \frac{1}{N} \sum_{n=1}^{N} K_n R(\hat{\vartheta}_R) K_n R(\hat{\vartheta}_R)'
\]

(3.20)

where the index \( n \) indicates the \( n \)-th component of the simulated score. Expressions (3.19) and (3.20) can be calculated by simulation.

4 Alternative approaches

4.1 Problems and alternative routes

In the previous two sections we discussed how a utility consistent labour supply model with taxes can be specified and estimated. It is clear that this procedure, which requires the imposition of coherency and the use of non-smooth simulators, is, even today, involved and nonstandard. It is tempting to look for simplifications of the model or the estimation procedure. With the model at hand, we can evaluate the 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. As indicated in section 2, 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 \textit{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 \textit{maximum maximorum} is in any of the convex sets. Finding the density of hours for each convex set is tedious but feasible (see section 2), 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.\footnote{This is made more precise in Appendix A.} This difficulty will arise in all but the simplest utility specifications.

Although the data will be described more fully below, it is worth observing that the problems discussed here are not academic. For the Dutch data that we will be using, budget sets faced by individuals may be highly non-convex and non-linear. Figure 4 shows a number of budget restrictions representative of the data we are using. The upper part shows nine different graphs for different levels of education and different percentiles of the before tax wage distribution. The benefit level is the sample mean of actual weekly benefits (see section 5). Non-labour income has been normalized to zero. To aid visual interpretation, Figure 5 shows one budget restriction, with auxiliary tangency lines drawn for different segments. One can see that the tangency lines imply substantial variation in virtual non-labor income (the intercept). To further illustrate the variation in marginal wage rates, Table 2 in section 5 provides values of marginal tax rates. These vary from zero to 72\%.

In view of the generally quite complicated nature of European social security and welfare systems, one may expect that certainly in Europe budget constraints will often be highly non-convex. A striking example for the U.K. has been constructed by Pudney (1989).

\begin{itemize}
  \item \textit{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 \(\epsilon_n, u_n, v_n\) 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.\footnote{This is also explained in detail in Appendix A. See the discussion leading up to (A.23) and the paragraph just below (A.23).} Since the
Figure 4: Illustrative budget sets based on data. Graph 1, 2, 3: lowest education level, graph 3, 4, 5: medium ed. lev., graph 7, 8, 9: highest education level, graph 1, 4, 7: 25 percentile wage distr., graph 2, 5, 8: median, graph 3, 6, 9: 75 perc. Currently (July 2003) a guilder is approximately equivalent to 0.50 US$.

Figure 5: Illustrative budget restrictions based on data.
budget constraint is the result of the interplay of the before tax 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\(^\text{16}\) 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 guaranteed to be 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.\(^\text{17}\)

- **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.

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

---

\(^{16}\) As stated above, the case of non-convex budget constraints necessitates the calculation of maximal utility on several convex sub-sets. In cases where such maxima correspond to kinks or corners, one needs to be able to calculate the direct utility associated with these points (or equivalently to find the virtual wage and non-labor income associated with these points). Unfortunately, various flexible forms satisfying curvature conditions globally have no known closed form for the direct utility function. See, for example, Diewert and Wales (1987, 1993, 1995), Lewbel (1995), Cooper and McLaren (1996).

\(^{17}\) In some special cases the need to guarantee coherency a priori does not arise. For instance, in a case where the budget set is convex and all random variation in hours for a given budget constraint is ascribed to heterogeneity of preferences, it is easy to see that the existence of observations at a kink will induce ML to impose coherency automatically, cf. MaCurdy, Green and Paarsch (1990). See also Blomquist (1995).
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.

- **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), Blomquist(1996)). 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.\(^{18}\)

- **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. See for example, Blomquist and Hansson-Brusewitz (1990), Bourguignon and Magnac (1990), Colombino and Del Boca (1990), Triest (1990), Van Soest, Woittiez, and Kapteyn (1990). 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. In addition, one has to be careful to deal properly with bunching at the kinks. For instance, Blundell, Duncan, and Meghir (1992, 1998) consider a case where the budget constraint has a long linear segment.

\(^{18}\) An alternative interpretation may be that the random term in the utility comparison represents optimization errors, which may be more natural in certain contexts.
They drop all observations that are not on this linear segment and instrument after tax wages. They observe, that in such a case bunching at or near a kink still leads to inconsistent estimates, in the same way in which running OLS in a Tobit model would yield inconsistent estimates. To avoid this problem the authors select “deep” into the linear segment and add a term to their estimation equation which represents the selectivity associated with that procedure.

Keane and Moffit (1998) estimate a structural model of labour supply with a nonlinear and nonconvex budget set. Their approach is in many respects different than the one we follow here. First, the emphasis is on the modeling of multiple program participation. To circumvent incoherency of the model due to possible nonconcavity of preferences, they discretize the choice set of hours by allowing only for the choice between ‘part-time’ hours and ‘full-time’ hours. This may be a legitimate restriction in a context in which the explanation of participation in welfare programs is the focal point of the analysis. Moreover, it is clear that this approach greatly reduces the problem of locating the worker’s optimum, since only two points on the budget set need to be compared for a given individual. However, observed variation in working hours and its interaction with the budget set, that is especially important for the analysis of female labour supply, is ignored. Second, Keane and Moffit (1998) compare the estimation of two methods, simulated maximum likelihood and the method of simulated moments, that are both consistent estimators under certain conditions. It is no surprise that the results obtained with these two methods do not deviate too much. Our approach is to compare popular methods that are commonly used, but which are in the present complicated model context no more than approximate methods. Third, Keane and Moffit (1998) also consider the estimation by MSS in which the wage are estimated in a first stage, as well as some variant of IV estimation. Due to the fact that they only use two levels of working hours and two points of the budget set the outcomes presumably will be less sensitive to the use of the method of estimation.

A paper that is more closely related to the current model is Van Soest (1995) who estimates a joint labor supply model of husbands and wives with a discretized budget set (six levels of weekly hours: 0, 10, ...,50 for both spouses), allowing for hours restrictions and random preferences. In contrast to his paper, one of our main goals in this papers is to compare a structural model of continuous labor supply with alternative simplified models or simplified estimation methods.

---

19 Bloemen (2002) set up a labour supply model, in which observed hours follows a discrete distribution due to demand side restrictions, but in which preferred hours still may take continuous values. In addition to observed, subjective information on desired hours is used to identify demand side restrictions from preferences.
4.2 Alternative methods of estimation

Following the discussion above, we will describe alternative methods of estimation. The methods we describe below either omit something from the original model or approximate certain parts of the model. This complicates the comparison of the outcomes of the different methods employed. But the alternative methods are tempting to use in practice because of their easier applicability.

4.2.1 Reduced form wage-participation model

Some of the approximate methods mentioned above, make use of predicted wages. Predicted wages are obtained in a first step. To estimate the parameters of the wage distribution, a reduced form wage-participation model is estimated. The joint wage-participation model is

\[
\begin{align*}
\ln w_n &= \eta' x_n + u_n \\
y_n^* &= \kappa' z_n + e_n \\
y_n^* &> 0 \text{ if working } (w_n \text{ observed}) \\
&\leq 0 \text{ if not } (w_n \text{ unobserved})
\end{align*}
\]

(4.1)

with

\[
\begin{pmatrix}
\ln \lambda_n \\
e_n
\end{pmatrix} \sim N \left( \begin{pmatrix} m_n \\ 0 \end{pmatrix}, \Sigma \right)
\]

(4.2)

and

\[
\Sigma = \begin{pmatrix} \tau^2 & \sigma_{ue} \\ \sigma_{ue} & 1 \end{pmatrix}
\]

(4.3)

where \( \lambda_n \) and \( m_n \) have been defined in (2.19) and (2.20). The variance of \( e_n \) has been normalized to one. The participation equation in (4.1) can be interpreted as an approximate reduced form of

\[
h_n^* = h(\exp(\eta' x_n + u_n), \mu_n; \alpha, q_n^* \omega + v_n) + e_n
\]

(4.4)

The vector of variables \( z_n \) should therefore contain all the variables included in the wage equation, as well as the variables appearing in the labor supply function \( h(.) \).

4.2.2 No random preferences, predicted budget constraints

The first approximate 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 (2.11) becomes

\[
h_n = h(w_n, \mu_n; \alpha, q_n^* \omega)
\]

(4.5)
Optimal labor supply is determined according to scheme (2.22) with \( v_n = 0 \). The only source of randomness now is measurement error \( \epsilon_n \) and the participation rule in this model is:  

\[
\begin{align*}
  h_n^* &= \tilde{h}_n + \epsilon_n \\
  h_n &= 0 \quad \text{if} \quad h_n^* \leq 0 \\
  &= h_n^* \quad \text{if} \quad h_n^* > 0
\end{align*}
\]  

(4.6)

The coherency constraint (2.18) remains the same.

One single predicted budget-constraint is used per observation, i.e., variation in the budget constraint due to variation in wages is ignored. Wages are predicted as the expected wage, based on the wage distribution defined in (2.19)-(2.20) using the estimates obtained by the estimation of the reduced form wage-participation model (4.1)-(4.3). Similarly, the prediction of the wage squared is determined as an expected value.  

### 4.2.3 IV, participants only

The next method we consider is instrumental variables. Only participants are taken into consideration and the budget constraint is linearized around the observed value of labor supply, \( h \). Suppose that \( H_{j-1} < h < H_j \); Then we are on the \( j \)-th segment and the budget constraint is approximated by the linear budget constraint with slope \( w_j \) and intercept \( \mu_j \) corresponding to segment \( j \). The equation to be estimated is:

\[
h = \alpha_1 + \alpha_2 \mu_j + \alpha_3 w_j + \frac{1}{2} \alpha_4 w_j^2 + q' \zeta + \epsilon
\]

(4.7)

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

Instrumental variables for the intercept \( \mu_j \), the slope \( w_j \) and its square \( w_j^2 \) have to be used. Obvious candidates are non-labor income \( \mu \), the gross wage rate \( w \) and its square, and individual characteristics appearing in the wage equation. For the gross wage to be a valid instrument, it has to be assumed that the gross wage rate and \( \epsilon \) are uncorrelated. Although we will impose this in the simulation, it is doubtful that the assumption will be true in practice. Hence, we present two sets of simulation results, for different sets of instruments. The first set does not include gross wage (or its square), but the second set does. In the IV-approach, two selectivity problems emerge. The first one is due to the fact that only participants are considered. To remedy this, we follow the traditional

---

20 Note that there is a difference between (4.6) and the model in (2.13) and (2.14). This difference arises due to the fact that \( \tilde{h}_n \) in (4.6) is not random since randomness in preferences is omitted.

21 The expected values of the wage and the squared wage have been determined by numerical integration.
approach of including an inverse Mills’ ratio as a correction term:

\[ h = \alpha_1 + \alpha_2 \mu_j + \alpha_3 w_j + \frac{1}{2} \alpha_4 w_j^2 + q' \zeta + \sigma_e \hat{\lambda} + \epsilon \] (4.8)

where \( \hat{\lambda} \) is the inverse Mills’ ratio estimated from (4.1). Wages are replaced by predicted wages, estimated by the reduced form wage-participation model.

The second selectivity problem stems from the linearization of the budget constraint around the observed number of hours (cf. Blundell, Duncan, Meghir (1998)). The model implies a bunching of observations around the various kinks in the budget constraint. This reduces the sensitivity of observed hours to wage differences across individuals and would lead to a downward bias of estimated wage elasticities, for instance. One solution, proposed by Blundell, Duncan, Meghir (1998) is to drop observations close to kink points and next to correct for the selectivity induced by this. In their application, where the budget constraint has only one kink, this is practically feasible, but in our case, with many kinks, it would amount to throwing away most of the observations. We have therefore ignored this second problem in the simulation. In itself this is consistent with the literature (see, e.g., Blomquist (1996))

4.2.4 MSS, predicted budget constraint

This method estimates the parameters of the utility function, but uses the predicted wages from the reduced form wage-participation model (4.1)-(4.3). So in the MSS procedure we only estimate the labour supply parameters and not the parameters associated with the wage equation. Keane and Moffit (1998) follow a similar approach, but apart from only using the predicted wages, they draw random errors from the wage distribution, which again is one step closer to the full estimation by MSS.

5 Empirical analysis

5.1 The data

The sample is a cross section data set from OSA.\textsuperscript{22} The sample used in the estimation of the model is a sample of 849 married females, drawn from the Dutch population in 1985. Table 1 contains some sample statistics. Most variable names are self explanatory. Five education levels are distinguished, where educ1 is the lowest education level and educ5 is the highest one. The latter category is used as the reference category in the analyses. The wage rate is obtained after dividing information on the wage income by the observed

\textsuperscript{22} “Organisatie voor Strategisch Arbeidsmarktonderzoek” (Organization for Strategic Labor Market Research); a government agency in the Netherlands.
number of working hours. This would introduce “division bias” in the estimation of the labour supply model, and therefore, we will allow for correlation between the measurement error of labour supply and the stochastics in the wage equation in the estimation of the structural model. Table 2 summarizes the tax system by listing the marginal tax rates of the ten different brackets. 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 welfare and 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. In the empirical analysis we have only accounted for one major non-convexity, arising from the high implicit tax on work for the unemployed.

Table 1: Sample statistics, Sample Size = 849

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours : weekly working hours</td>
<td>27.3</td>
<td>14.5</td>
</tr>
<tr>
<td>Gross Wage Rate : guilders/hour</td>
<td>14.5</td>
<td>5.9</td>
</tr>
<tr>
<td>Family size : number of persons</td>
<td>3.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Age</td>
<td>37.2</td>
<td>9.9</td>
</tr>
<tr>
<td>Non-labor income (guilders/week)</td>
<td>719.1</td>
<td>308.3</td>
</tr>
<tr>
<td>Benefit level (guilders/week): positive values only</td>
<td>159.5</td>
<td>58.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor market state: Number of employed</td>
<td>331</td>
<td>40.0</td>
</tr>
<tr>
<td>Educ1</td>
<td>217</td>
<td>25.6</td>
</tr>
<tr>
<td>Educ2</td>
<td>223</td>
<td>26.3</td>
</tr>
<tr>
<td>Educ3</td>
<td>322</td>
<td>37.9</td>
</tr>
<tr>
<td>Educ4</td>
<td>71</td>
<td>8.3</td>
</tr>
<tr>
<td>Educ5</td>
<td>16</td>
<td>1.9</td>
</tr>
<tr>
<td>Presence of children with age below 6</td>
<td>221</td>
<td>26.0</td>
</tr>
</tbody>
</table>
Table 2: Characterization of the tax system

<table>
<thead>
<tr>
<th>bracket</th>
<th>income bracket (guilders/week)</th>
<th>marginal tax rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt; 145</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>146-273</td>
<td>0.16</td>
</tr>
<tr>
<td>3</td>
<td>274-348</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>349-502</td>
<td>0.32</td>
</tr>
<tr>
<td>5</td>
<td>503-569</td>
<td>0.42</td>
</tr>
<tr>
<td>6</td>
<td>570-655</td>
<td>0.52</td>
</tr>
<tr>
<td>7</td>
<td>656-724</td>
<td>0.67</td>
</tr>
<tr>
<td>8</td>
<td>725-793</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>794-1331</td>
<td>0.70</td>
</tr>
<tr>
<td>10</td>
<td>&gt; 1331</td>
<td>0.72</td>
</tr>
</tbody>
</table>

5.2 Monte Carlo results

In this section we analyse the performance of our MSS estimator, described in section 3, and the various alternatives, addressed in section 4.2.

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 of family size (with parameter $\zeta_1$) and a dummy indicator taking the value one if the woman has children below 6, and zero if not (parameter $\zeta_2$). In the Monte Carlo experiment, the variables in the wage equation are a constant term (parameter $\eta_1$), log-age (parameter $\eta_2$) and log-age squared (parameter $\eta_3$). To restrict the number of parameters in the Monte Carlo experiment we have omitted dummy indicators for the level of education.\(^{23}\) The parameters of the negative gamma distribution are $\gamma_1$ and $\gamma_2$. 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 3 and the decision rules in (2.13) and (2.14). 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 model. Benefits are measured in guilders per week.

We use the simplex method by Nelder and Meade (1965) (see Press et al. (1986) for a survey) locate the MSS estimate. The algorithm starts with the initialization of the ‘simplex’: this is a series of $s+1$ different starting value vectors for the parameter vector,

\(^{23}\) These dummies are included in the estimation on the real data.
with $s$ the dimension of the parameter space. For practical purpose, we may use the parameter values of the model without random preferences, the IV method and the reduced form wage model as one of the initial parameter values. For the other parameter vectors, we construct values that deviate from these values in all of the different dimensions of the parameter space: the larger the deviation, the larger is the initial simplex and the larger is the parameter space that is considered by the algorithm, but also the longer it takes for the algorithm to converge. This way of initializing the starting simplex allows the algorithm to search over a large space.

For the MSS variants we use $R = 10$ and $R_Z = 800$ replications per individual to simulate the score. The value of $\tilde{w}$ in (2.20) has been chosen to be 5% higher than the largest wage observed in the (Monte Carlo or real) sample. The value $\hat{w}$ in (2.18) is set to the virtual wage associated with the highest tax bracket if the wage is $\tilde{w}$, and $\hat{\mu}$ is the virtual income based on the largest value of $\mu$ in the sample.

Twenty Monte Carlo datasets have been generated, by the completely specified model. The limited number of Monte Carlo experiments is dictated by the heavy computational burden of some of the estimation methods. Table 3 show the results. For each method, the table provides the mean of the estimates over the Monte Carlo datasets, the standard deviation of the estimates, the mean of the estimated standard errors, and the absolute value of the relative error, i.e., $|\hat{\theta} - \theta_0|/|\theta_0|$.

The results show that in general MSS performs better than the other methods, and often substantially so. Of course this is not surprising, since MSS is the only estimation method exploiting the correct data generating process. Thus the results mainly illustrate the weaknesses of the various simplifications represented by the other estimation methods. The reported standard deviations and mean standard errors of MSS are usually of similar magnitude, as they should be. IV which does not use gross wages as instruments performs worst. If the gross wage is used as instrument, IV performs better. Recall however that we have imposed the assumption that gross wages are not correlated with the error in (4.7), which seems unlikely in reality.

In the Monte Carlo experiment for MSS, the true parameter value has been used to construct the vector of instruments $Z$. In reality, we need estimated parameter values to construct the matrix of instruments $Z$. Therefore we also performed an experiment in which we use estimated values for the parameters. The estimation results from the model without random preferences were used as values for the utility parameters, whereas the estimation results of the wage-participation model were used to get values for the parameters of the wage distribution. The model has been estimated by MSS, with instruments based on these values. The estimates were used to construct a new matrix
Z, and a second estimation round followed. The results of this procedure are in the final column of table 3. The main difference between MSS and MSS using estimated instruments seems to be the larger variation in the estimates of the different Monte Carlo data sets, as well as the larger standard errors.

Table 4 presents mean elasticities for the different estimation methods. Elasticities are computed by simulating a model twice for each parameter set. In the first run hours are generated for all observations. In the second run all wages are increased by 5% and hours are generated again. The percentage change in mean hours and mean participation then yields the elasticities we are looking for. The numbers in parentheses are standard deviations of the computed elasticities across the twenty Monte Carlo datasets. Table 4 shows that the elasticities are quite sensitive to the estimation strategy adopted. Interestingly, the elasticities corresponding to “no random preferences” are even closer to the elasticities computed with the true parameter values than those obtained with MSS. The elasticities obtained with MSS with an estimated instrument matrix show a larger deviation from the elasticities based on the true values. The other two methods generate elasticities that are quite different from the “true” elasticities.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>true value</th>
<th>no rand. pref.</th>
<th>IV</th>
<th>IV gross wages</th>
<th>MSS, n-st.bc</th>
<th>MSS</th>
<th>MSS, Z estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_1 ) (const)</td>
<td>mean</td>
<td>14.1</td>
<td>3.5</td>
<td>34.4</td>
<td>27.2</td>
<td>3.4</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>10.9</td>
<td>42.8</td>
<td>15.8</td>
<td>11.6</td>
<td>8.8</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td>st. err.</td>
<td>31.9</td>
<td>46.0</td>
<td>11.8</td>
<td>4.9</td>
<td>4.6</td>
<td>7.4</td>
</tr>
<tr>
<td></td>
<td>rel. err.</td>
<td>0.75</td>
<td>1.5</td>
<td>0.93</td>
<td>0.76</td>
<td>0.023</td>
<td>0.029</td>
</tr>
<tr>
<td>( \alpha_2 ) (non-lab.inc)</td>
<td>mean</td>
<td>-0.047</td>
<td>-0.037</td>
<td>-0.014</td>
<td>-0.036</td>
<td>-0.027</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.010</td>
<td>0.015</td>
<td>0.012</td>
<td>0.011</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>st. err.</td>
<td>0.012</td>
<td>0.022</td>
<td>0.011</td>
<td>0.022</td>
<td>0.003</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>rel. err.</td>
<td>0.22</td>
<td>0.71</td>
<td>0.23</td>
<td>0.42</td>
<td>0.08</td>
<td>0.26</td>
</tr>
<tr>
<td>( \alpha_3 ) (wage)</td>
<td>mean</td>
<td>10.69</td>
<td>3.2</td>
<td>3.2</td>
<td>7.23</td>
<td>7.1</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.6</td>
<td>12.2</td>
<td>3.2</td>
<td>2.8</td>
<td>1.2</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>st. err.</td>
<td>8.2</td>
<td>12.9</td>
<td>2.5</td>
<td>3.3</td>
<td>1.0</td>
<td>7.8</td>
</tr>
<tr>
<td></td>
<td>rel. err.</td>
<td>0.37</td>
<td>0.70</td>
<td>0.32</td>
<td>0.34</td>
<td>0.03</td>
<td>0.14</td>
</tr>
<tr>
<td>( \alpha_4 ) (.5 wage sqr)</td>
<td>mean</td>
<td>-0.26</td>
<td>-0.043</td>
<td>-0.13</td>
<td>-0.16</td>
<td>-0.10</td>
<td>-0.17</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.18</td>
<td>1.1</td>
<td>0.32</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>st. err.</td>
<td>0.69</td>
<td>1.1</td>
<td>0.24</td>
<td>0.57</td>
<td>0.21</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>rel. err.</td>
<td>0.37</td>
<td>0.51</td>
<td>0.41</td>
<td>0.62</td>
<td>0.32</td>
<td>0.33</td>
</tr>
<tr>
<td>( \zeta_1 ) (log(family size))</td>
<td>mean</td>
<td>-24</td>
<td>-16.0</td>
<td>-6.6</td>
<td>-18.3</td>
<td>-13.9</td>
<td>-25.4</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>15.6</td>
<td>6.7</td>
<td>6.8</td>
<td>12.4</td>
<td>8.0</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td>st. err.</td>
<td>6.8</td>
<td>12.1</td>
<td>5.4</td>
<td>9.2</td>
<td>5.8</td>
<td>12.5</td>
</tr>
<tr>
<td></td>
<td>rel. err.</td>
<td>0.83</td>
<td>0.73</td>
<td>0.24</td>
<td>0.42</td>
<td>0.06</td>
<td>0.28</td>
</tr>
<tr>
<td>( \zeta_2 ) (dum.child.age &lt;6)</td>
<td>mean</td>
<td>-13.9</td>
<td>-14.0</td>
<td>-3.9</td>
<td>-10.2</td>
<td>-11.7</td>
<td>-14.8</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>13.6</td>
<td>5.6</td>
<td>3.5</td>
<td>9.9</td>
<td>4.1</td>
<td>17.7</td>
</tr>
<tr>
<td></td>
<td>st. err.</td>
<td>7.9</td>
<td>6.6</td>
<td>3.8</td>
<td>8.0</td>
<td>4.1</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td>rel. err.</td>
<td>0.05</td>
<td>0.72</td>
<td>0.27</td>
<td>0.16</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>( \sigma_\epsilon )</td>
<td>mean</td>
<td>19.3</td>
<td>33.2</td>
<td>23.1</td>
<td>28.7</td>
<td>34.3</td>
<td>24.0</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.2</td>
<td>4.5</td>
<td>2.7</td>
<td>4.6</td>
<td>5.5</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>st. err.</td>
<td>1.3</td>
<td>—</td>
<td>—</td>
<td>1.1</td>
<td>2.5</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td>rel. err.</td>
<td>0.33</td>
<td>0.19</td>
<td>0.48</td>
<td>0.78</td>
<td>0.24</td>
<td>0.28</td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>mean</td>
<td>3.0</td>
<td>—</td>
<td>—</td>
<td>3.2</td>
<td>3.2</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.9</td>
<td>1.1</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>st. err.</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.23</td>
<td>0.4</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>rel. err.</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.05</td>
<td>0.07</td>
<td>0.35</td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>mean</td>
<td>4.0</td>
<td>—</td>
<td>—</td>
<td>3.96</td>
<td>4.7</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.29</td>
<td>2.2</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>st. err.</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.4</td>
<td>1.1</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>rel. err.</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.01</td>
<td>0.19</td>
<td>0.08</td>
</tr>
</tbody>
</table>

**parameters of wage equation**

| \( \eta_1 \) (const.)         | mean       | 1.4           | —     | —             | —             | —     | 1.4             |
|                               | SD         | —              | —     | —             | 0.37          | 0.47  | 0.41            |
|                               | st. err.   | —              | —     | —             | 0.20          | 0.41  | 0.04            |
|                               | rel. err.  | —              | —     | —             | 0.04          | 0.70  | 0.70            |
| \( \eta_2 \) log(age/17)      | mean       | 2.75          | —     | —             | —             | —     | 3.4             |
|                               | SD         | —              | —     | —             | 0.5           | 1.1   | 1.1             |
|                               | st. err.   | —              | —     | —             | 0.4           | 0.6   | 0.6             |
|                               | rel. err.  | —              | —     | —             | 0.23          | 0.24  | 0.24            |
| \( \eta_3 \) log(age/17)sq    | mean       | -2.2          | —     | —             | —             | —     | -3.1            |
|                               | SD         | —              | —     | —             | 0.81          | 0.68  | 0.68            |
|                               | st. err.   | —              | —     | —             | 0.33          | 0.38  | 0.07            |
|                               | rel. err.  | —              | —     | —             | 0.43          | 0.07  | 0.07            |
| \( \tau \)                    | mean       | 0.981         | —     | —             | —             | —     | 1.082           |
|                               | SD         | —              | —     | —             | 0.103         | 0.251 | 0.251           |
|                               | st. err.   | —              | —     | —             | 0.02          | 0.03  | 0.03            |
|                               | rel. err.  | —              | —     | —             | 0.10          | 0.14  | 0.14            |
Table 4: Elasticities in Monte Carlo experiments
(standard deviations in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>true values</th>
<th>no rand. pref.</th>
<th>IV</th>
<th>MSS, n-st. bdg.</th>
<th>MSS,</th>
<th>MSS, Z estimated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wage elasticity of hours</td>
<td>0.71</td>
<td>0.71</td>
<td>0.20</td>
<td>0.85</td>
<td>0.69</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.13)</td>
<td>(0.36)</td>
<td>(0.35)</td>
<td>(0.11)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Wage elasticity of Participation</td>
<td>0.36</td>
<td>0.34</td>
<td>0.07</td>
<td>0.62</td>
<td>0.40</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.06)</td>
<td>(0.14)</td>
<td>(0.42)</td>
<td>(0.11)</td>
<td>(0.12)</td>
</tr>
</tbody>
</table>

5.3 Empirical results

Let us now turn to the real data. For each method of estimation, two variants are estimated, which differ in the way the budget constraint has been constructed: For almost all the working females, benefits are observed to be zero. Consequently, the first variant constructs the budget constraint with zero benefits for these women and hence the budget set is convex. The second variant assumes that working women may obtain an unemployment benefit if they leave employment. As a consequence, the budget constraint for working women becomes non-convex as well, and this will affect their decision to work or not. The level of the unemployment benefit is about 70% of net earnings. Not every woman will receive an unemployment benefit. This typically depends on work history, and whether or not she quits or is laid off. To account for this in a rough way, we assume that the women anticipate to get 50% of the unemployment benefit level. Thus, the first variant is based on the assumption that females who work ignore the possibility to receive benefits by leaving paid employment, whereas the second variant assumes that everybody does to some extent.

Thus far, the only constraint faced by an individual is the budget constraint. To allow for other restrictions, we estimate an extension of the structural model with MSS, and in the tables this method is referred to as MSS plus. Similar to an approach taken by Blundell, Ham, and Meghir (1987), we assume that there is an additional mechanism that determines the probability of employment given participation:

\[ y^*_n = \kappa' z_n + sv_n + \epsilon_n \]  

(5.1)

in which \( v_n \) is the random preference variable that has been introduced in (2.9). We call this the employment equation. The employment equation is added to the labor supply equation (2.12) and the wage specification in (2.19)-(2.20). The parameter \( s \) represents a possible dependence between preferences and the likelihood of being employed. The
following distributional assumption is made:

\[ \begin{pmatrix} e_n \\ u_n \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{ee}^2 & \sigma_{eu} \\ \sigma_{ue} & \sigma_{uu} \end{pmatrix} \right) \]

(5.2)

The participation decision (2.13)-(2.14) is now extended to:

\[ h_n = h_n^* \quad \text{if} \quad h_n^* > 0 \quad \text{and} \quad h_n(w_n, \mu_n; \alpha, q_n^t \omega + v_n) > 0 \quad \text{and} \quad y_n^* > 0 \]
\[ h_n = 0 \quad \text{if} \quad h_n^* \leq 0 \quad \text{or} \quad h_n(w_n, \mu_n; \alpha, q_n^t \omega + v_n) \leq 0 \quad \text{or} \quad y_n^* \leq 0 \]

(5.3) (5.4)

Thus, we will observe a female working if she wants to work, if she can find employment and if measurement errors do not hide this fact from us. The structure of (5.1) is almost identical to the participation equation in (4.1). The latter equation can be seen as a reduced form reflecting both demand and supply factors determining whether a female will work or not. Due to the presence of (2.13) and (2.14), (5.1) can be interpreted with less ambiguity as an employment equation, rather than a participation equation.

Further, note that in the estimation by MSS, we also allow for correlation between wages and labor supply, by allowing for a nonzero covariance \( \sigma_{eu} \) between \( e_n \) and \( u_n \).

In all of the methods employing the method of simulated scores, the number of drawings \( R \) to simulate probabilities is equal to \( R = 30 \). The number of drawings to compute the matrix of instruments is \( R_Z = 2000 \).

For the estimation by IV, the choice of instruments has been based on preliminary regressions, in which endogenous regressors have been regressed on (potential) instruments. The following instruments have been included: log of family size, dummy for children with age below six, log age, log age squared, educ1, educ2, educ3 and educ4. Of these instruments, only the education variables are not included directly in (4.7).

The first column of table 5 presents the estimation results of the participation equation in (4.1), estimated jointly with the wage equation. The parameter estimates of the corresponding wage equation are presented in the first column of table 7. These estimates are used to construct predicted wages for some of the methods of estimation. In the estimation of the participation equation, some squared and cross terms have been included, as the participation equation is to approximate the participation decision of the structural model. We actually have tried more cross and squared terms than are presented here, but on the basis of likelihood ratio tests we found those not to be significant.

Table 6 presents the estimation results for the preference parameters for both variants, i.e., with zero benefits for working women (variant 1) and with benefits of working women calculated as half times 70% of their net earnings (variant 2). For none of the estimation

\[ \sigma_e^2 \] is normalized to one.
methods the quadratic wage term is significant. For all of the methods, except for IV, family size has a negative effect on labor supply. This negative effect is not significant for MSS with a non-stochastic budget constraint and MSS plus. Note that for the latter method family size also enters the employment equation (table 5) with a negative sign (significant at the 10% level for variant 1). The dummy for the presence of children with age below six, is negative for the method without random preferences (but significant for variant 1 only), IV (not significant), and MSS (both variants significant). For MSS plus we see a positive sign, and an insignificantly negative sign in the employment equation (table 5). For similar reasons it is hard to interpret the outcomes for the parameters determining the effect of age. Also here these interact in complicated ways with the employment equation.

All of the methods, except IV, have been implemented imposing the coherency constraint. In the estimated parameter values, the constraint is not binding. The estimates obtained by IV do not satisfy the coherency constraint.

The predominant impression is that the estimation outcomes vary considerably across both estimation methods and across model assumptions (in particular whether working females could exit the labor force and collect benefits or not).

Table 7 shows the estimates of the parameters of the wage equation. We see that, except for the intercept, the signs of all estimated parameters are identical across the various columns. Yet, the magnitude of the parameters vary quite a bit. For instance, the outcomes for the education coefficients in different columns tell us rather different stories regarding the effect of education on wages. In particular MSS plus suggests lower returns to education than MSS.

Table 8 presents wage elasticities of hours and participation, computed similarly to the ones in Table 4. In the simulations we now also take the employment equation into account, whenever appropriate. The elasticities vary substantially, both across estimation methods and across variants. Note that the elasticities for IV may be hard to interpret since the parameters do not satisfy the coherency constraint: the participation elasticity is negative.

To gain more insight in the effects of the models for individuals, we have calculated elasticities on an individual basis as well. We simulated for each individual 1000 drawings from the distribution of the errors in the model, and calculated for each individual the impact a 5% wage increase. Thus we can determine the distribution of the elasticities across individuals. Table 9 lists the mean, minimum and maximum, and the 25%, 50%, and 75% quantiles of the distribution of the participation elasticities. The elasticities for the method without random preferences are zero for a large part of the individuals.
The fact that the model does not allow for random variation in preferences and that the budget constraint is non-random probably results in many individuals being stuck at kinks and corners. Note that the lowest participation elasticity is negative. This is a strange result at first sight, since the model satisfies coherency. We found that for individuals with a negative participation elasticity the following is going on: (i) optimal labor supply is positive, both before and after the 5% wage increase, (ii) the optimal level of labor supply is lower after the wage increase than before the wage increase, (iii) adding measurement error results in a positive number of working hours before the wage increase and zero working hours after. Thus optimal labor supply meets the coherency condition, but for observed labor supply, measurement error leads to an observed negative participation elasticity.

For IV we mostly see zero, and some negative values for the participation elasticities. For MSS with a non-stochastic budget constraint, and also for MSS, we see that the median participation elasticities of variant 1 and variant 2 are the same. Differences in the mean value are caused by values of the elasticities in the tails of the distribution.

Table 10 shows some features of the distribution of the hours elasticities. For the method without random preferences and the method with a non-stochastic budget constraint, there is not much difference between the median values of the elasticities, between variant 1 and variant 2. The difference in the median elasticities for MSS and MSS plus between variant 1 and 2 show the same direction as we already found in table 8. This shows that the more flexible specifications lead to larger differences in outcomes due to a change in the specification of the budget constraint due to non-convexities.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Reduced form</th>
<th>MSS plus 1st v.</th>
<th>MSS plus 2nd v.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa_1$ (const)</td>
<td>4.5</td>
<td>4.8</td>
<td>4.5</td>
</tr>
<tr>
<td>$\kappa_2$ (log(fam. size))</td>
<td>(0.8)</td>
<td>(3.3)</td>
<td>(2.5)</td>
</tr>
<tr>
<td>$\kappa_3$ (d. child. &lt; 6)</td>
<td>-3.6</td>
<td>-4.5</td>
<td>-4.7</td>
</tr>
<tr>
<td>$\kappa_4$ (non-labor income)</td>
<td>(0.8)</td>
<td>(2.5)</td>
<td>(2.9)</td>
</tr>
<tr>
<td>$\kappa_5$ (log(age/17))</td>
<td>1.5</td>
<td>1.7</td>
<td>1.6</td>
</tr>
<tr>
<td>$\kappa_6$ (square of log(age/17))</td>
<td>-4.2</td>
<td>-4.3</td>
<td>-4.2</td>
</tr>
<tr>
<td>$\kappa_7$ (educ1)</td>
<td>-1.7</td>
<td>-1.9</td>
<td>-1.9</td>
</tr>
<tr>
<td>$\kappa_8$ (educ2)</td>
<td>-1.9</td>
<td>-1.9</td>
<td>-1.9</td>
</tr>
<tr>
<td>$\kappa_9$ (educ3)</td>
<td>-1.5</td>
<td>-0.8</td>
<td>-0.5</td>
</tr>
<tr>
<td>$\kappa_{10}$ (educ4)</td>
<td>-0.93</td>
<td>0.36</td>
<td>0.12</td>
</tr>
<tr>
<td>$\kappa_{11}$ (log(fs)*log(age/17))</td>
<td>3.9</td>
<td>4.1</td>
<td>4.0</td>
</tr>
<tr>
<td>$\kappa_{12}$ (log(fs)*$\mu$)</td>
<td>-17.0</td>
<td>-17.7</td>
<td>-17.1</td>
</tr>
<tr>
<td>$\kappa_{13}$ ($\mu^2$)</td>
<td>92.0</td>
<td>93.5</td>
<td>93.7</td>
</tr>
<tr>
<td>$\sigma_{eu}$ (employment-wage)</td>
<td>-0.02</td>
<td>-0.005</td>
<td>0.03</td>
</tr>
<tr>
<td>$\sigma_{ex}$ (employment-hours)</td>
<td>—</td>
<td>0.52</td>
<td>0.50</td>
</tr>
<tr>
<td>$s$ (employment-rand. pref.)</td>
<td>—</td>
<td>-0.028</td>
<td>-0.020</td>
</tr>
</tbody>
</table>

$R = 30, R^2 = 2000$ for MSS plus; standard errors in parentheses

1st v. = first variant, 2nd v. = second variant
Table 6: Preference parameters

<table>
<thead>
<tr>
<th>parameter</th>
<th>No random preferences</th>
<th>IV</th>
<th>MSS, non-stoch. budget constraint</th>
<th>MSS</th>
<th>MSS plus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st v.</td>
<td>2nd v.</td>
<td>1st &amp; 2nd v.</td>
<td>1st v.</td>
<td>2nd v.</td>
</tr>
<tr>
<td>$\alpha_1$ (const)</td>
<td>34.3</td>
<td>-23.9</td>
<td>31.7</td>
<td>5.1</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>(15.6)</td>
<td>(12.9)</td>
<td>(23.3)</td>
<td>(4.6)</td>
<td>(4.2)</td>
</tr>
<tr>
<td>$\alpha_2$ (nlab. inc.)</td>
<td>-0.0096</td>
<td>-0.00021</td>
<td>-0.0020</td>
<td>-0.0029</td>
<td>-0.0031</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.0020)</td>
<td>(0.051)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>$\alpha_3$ (wage)</td>
<td>3.62</td>
<td>38.8</td>
<td>2.88</td>
<td>3.7</td>
<td>3.6</td>
</tr>
<tr>
<td>$\alpha_4$ (0.5 sq. w.)</td>
<td>(2.48)</td>
<td>(6.5)</td>
<td>(4.32)</td>
<td>(3.4)</td>
<td>(2.9)</td>
</tr>
<tr>
<td></td>
<td>(0.18)</td>
<td>(2.6)</td>
<td>(0.34)</td>
<td>(0.72)</td>
<td>(0.46)</td>
</tr>
<tr>
<td>$\zeta_1$ (log(fsize))</td>
<td>-56.9</td>
<td>-9.7</td>
<td>5.30</td>
<td>-23.2</td>
<td>-23.0</td>
</tr>
<tr>
<td></td>
<td>(11.6)</td>
<td>(2.5)</td>
<td>(2.38)</td>
<td>(17.4)</td>
<td>(15.3)</td>
</tr>
<tr>
<td>$\zeta_2$ (d. ch. &lt; 6)</td>
<td>-16.8</td>
<td>-2.02</td>
<td>-2.31</td>
<td>11.0</td>
<td>11.7</td>
</tr>
<tr>
<td></td>
<td>(4.8)</td>
<td>(2.3)</td>
<td>(3.14)</td>
<td>(8.7)</td>
<td>(8.5)</td>
</tr>
<tr>
<td>$\zeta_3$ (log(age/17))</td>
<td>25.7</td>
<td>-47.4</td>
<td>-20.2</td>
<td>87.4</td>
<td>87.5</td>
</tr>
<tr>
<td></td>
<td>(24.6)</td>
<td>(16.7)</td>
<td>(16.9)</td>
<td>(54.0)</td>
<td>(44.5)</td>
</tr>
<tr>
<td>$\zeta_4$ (Lage/17 sq.)</td>
<td>-49.6</td>
<td>21.2</td>
<td>9.5</td>
<td>-64.7</td>
<td>-65.0</td>
</tr>
<tr>
<td></td>
<td>(17.1)</td>
<td>(10.3)</td>
<td>(11.3)</td>
<td>(41.5)</td>
<td>(35.6)</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(0.97)</td>
<td>(0.94)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>209.2</td>
<td>207.3</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(164.2)</td>
<td>(149.3)</td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>19.2</td>
<td>11.8</td>
<td>9.7</td>
<td>36.7</td>
<td>37.2</td>
</tr>
<tr>
<td></td>
<td>(0.9)</td>
<td>(0.5)</td>
<td>-</td>
<td>(18.0)</td>
<td>(14.6)</td>
</tr>
<tr>
<td>$\sigma_{\epsilon u}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(0.5)</td>
<td>(2.4)</td>
</tr>
</tbody>
</table>

$R = 30, R^2 = 2000$ for MSS variants; standard errors in parentheses;
1st v. = first variant, 2nd v. = second variant

Table 7: Parameters of the wage equation

<table>
<thead>
<tr>
<th>parameter</th>
<th>reduced form</th>
<th>MSS</th>
<th>MSS plus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st v.</td>
<td>2nd v.</td>
<td>1st v.</td>
</tr>
<tr>
<td>$\eta_1$ (const)</td>
<td>2.5</td>
<td>1.8</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td>(0.53)</td>
<td>(0.33)</td>
</tr>
<tr>
<td>$\eta_2$ (log(age/17))</td>
<td>1.9</td>
<td>2.6</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
<td>(0.66)</td>
<td>(0.94)</td>
</tr>
<tr>
<td>$\eta_3$ (square of log(age/17))</td>
<td>-1.3</td>
<td>-1.1</td>
<td>-0.79</td>
</tr>
<tr>
<td></td>
<td>(0.3)</td>
<td>(0.39)</td>
<td>(0.51)</td>
</tr>
<tr>
<td>$\eta_4$ (educ1)</td>
<td>-0.64</td>
<td>-1.08</td>
<td>-1.66</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.38)</td>
<td>(0.78)</td>
</tr>
<tr>
<td>$\eta_5$ (educ2)</td>
<td>-0.53</td>
<td>-1.25</td>
<td>-1.00</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.40)</td>
<td>(0.55)</td>
</tr>
<tr>
<td>$\eta_6$ (educ3)</td>
<td>-0.45</td>
<td>-0.79</td>
<td>-0.85</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.33)</td>
<td>(0.48)</td>
</tr>
<tr>
<td>$\eta_7$ (educ4)</td>
<td>-0.21</td>
<td>-0.58</td>
<td>-0.67</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.27)</td>
<td>(0.45)</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.47</td>
<td>0.86</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.13)</td>
<td>(0.09)</td>
</tr>
</tbody>
</table>

$R = 30, R^2 = 2000$ for MSS variants; standard errors in parentheses;
1st v. = first variant, 2nd v. = second variant
Table 8: Elasticities

<table>
<thead>
<tr>
<th>Method</th>
<th>Variant</th>
<th>mean</th>
<th>min</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>No random preferences</td>
<td>1</td>
<td>0.26</td>
<td>-1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4.6</td>
</tr>
<tr>
<td>IV</td>
<td>1</td>
<td>-0.02</td>
<td>-5.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MSS non stochastic budget c.</td>
<td>1</td>
<td>0.15</td>
<td>-0.48</td>
<td>0</td>
<td>0.18</td>
<td>0.26</td>
<td>4.5</td>
</tr>
<tr>
<td>MSS</td>
<td>1</td>
<td>0.14</td>
<td>-0.26</td>
<td>0.06</td>
<td>0.12</td>
<td>0.20</td>
<td>1.2</td>
</tr>
<tr>
<td>MSS plus</td>
<td>1</td>
<td>0.90</td>
<td>-1.8</td>
<td>0.11</td>
<td>0.29</td>
<td>0.51</td>
<td>249</td>
</tr>
<tr>
<td>No random preferences</td>
<td>2</td>
<td>0.07</td>
<td>-6.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.9</td>
</tr>
<tr>
<td>IV</td>
<td>2</td>
<td>-0.01</td>
<td>-3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MSS non stochastic budget c.</td>
<td>2</td>
<td>7.1</td>
<td>-5.3</td>
<td>0.02</td>
<td>0.18</td>
<td>0.27</td>
<td>2900</td>
</tr>
<tr>
<td>MSS</td>
<td>2</td>
<td>0.35</td>
<td>-19</td>
<td>0</td>
<td>0.12</td>
<td>0.37</td>
<td>38</td>
</tr>
<tr>
<td>MSS plus</td>
<td>2</td>
<td>4.4</td>
<td>-19</td>
<td>0</td>
<td>0</td>
<td>0.13</td>
<td>662</td>
</tr>
</tbody>
</table>

standard errors in parentheses; 1st v. = first variant, 2nd v. = second variant

Table 9: The distributions of the participation elasticities

<table>
<thead>
<tr>
<th>Method</th>
<th>Variant</th>
<th>mean</th>
<th>min</th>
<th>25%</th>
<th>median</th>
<th>75%</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>No random preferences</td>
<td>1</td>
<td>0.51</td>
<td>-0.94</td>
<td>0</td>
<td>0</td>
<td>0.97</td>
<td>6.0</td>
</tr>
<tr>
<td>IV</td>
<td>1</td>
<td>-0.07</td>
<td>-14</td>
<td>-0.06</td>
<td>0.03</td>
<td>0.10</td>
<td>0.56</td>
</tr>
<tr>
<td>MSS non stochastic budget c.</td>
<td>1</td>
<td>0.40</td>
<td>-0.86</td>
<td>0.04</td>
<td>0.57</td>
<td>0.69</td>
<td>13</td>
</tr>
<tr>
<td>MSS</td>
<td>1</td>
<td>0.36</td>
<td>-0.88</td>
<td>0.27</td>
<td>0.39</td>
<td>0.60</td>
<td>2.4</td>
</tr>
<tr>
<td>MSS plus</td>
<td>1</td>
<td>1.4</td>
<td>-2.3</td>
<td>0.68</td>
<td>0.90</td>
<td>1.26</td>
<td>235</td>
</tr>
<tr>
<td>No random preferences</td>
<td>2</td>
<td>0.38</td>
<td>-14</td>
<td>0</td>
<td>0</td>
<td>0.81</td>
<td>5.6</td>
</tr>
<tr>
<td>IV</td>
<td>2</td>
<td>-0.02</td>
<td>-14</td>
<td>-0.06</td>
<td>0.03</td>
<td>0.11</td>
<td>23</td>
</tr>
<tr>
<td>MSS non stochastic budget c.</td>
<td>2</td>
<td>6.3</td>
<td>-9.3</td>
<td>0.10</td>
<td>0.58</td>
<td>0.69</td>
<td>1899</td>
</tr>
<tr>
<td>MSS</td>
<td>2</td>
<td>0.59</td>
<td>-19</td>
<td>0.33</td>
<td>0.52</td>
<td>0.67</td>
<td>36</td>
</tr>
<tr>
<td>MSS plus</td>
<td>2</td>
<td>11.7</td>
<td>-20</td>
<td>0.27</td>
<td>0.41</td>
<td>0.91</td>
<td>1603</td>
</tr>
</tbody>
</table>
6 Conclusions

Both the Monte Carlo results and the estimation results for real data show a large variation of outcomes across estimation methods. Furthermore, the outcomes are quite sensitive to the specification of the budget constraint (in our case, the entitlement to unemployment benefits if a working female would quit her job). For the Monte Carlo experiments 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. This sensitivity of results to economic and statistical assumptions is reminiscent of the outcomes reported by Mroz (1987), although we concentrate on different aspects than his paper.

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.

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 stylized environment and then concentrated on a utility consistent specification of 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 and of the budget constraint facing the individual, we would anticipate even more relevance of such treatment in more complicated environments.

The approach taken is in principle quite general. For instance, the Hausman-Ruud specification can be replaced by any other flexible specification that permits the calculation of direct utility of given leisure consumption combinations. Extensions of the decision problem, for instance by adding further complications to the budget set (e.g., due to demand side restrictions) can be accommodated in the same framework without great difficulty.
A Appendix. Computational details

In this appendix the computational details of the simulation of the score are described. 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 \( \mu_{nj} \) denote the intercept of the \( j \)-th segment of the budget constraint, as indicated in figure 3, where \( j = 1, \ldots, m \). 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 \( w_{nj} \), \( w_{n0} < w_{n1}, w_{nj} > w_{nj+1}, j = 1, \ldots, m - 1 \), and \( H_{nj} \) is the kink point between the \( j \)-th and \( (j+1) \)-th segment, \( j = 0, \ldots, m - 1 \). A convex budget set without the social security appears as a special case if \( H_{n0} = 0 \). If we allow for variation in the gross wage \( w_n \), the slopes \( w_{nj} \) and the kink points \( H_{nj} \) will depend \( w_n \). To express this dependence, we may write:

\[
\begin{align*}
  w_{nj} &= w_j(w_n) \quad (A.1) \\
  H_{nj} &= H_j(w_n) \quad (A.2) \\
  w'_j(w_n) &= 0 \quad (A.3) \\
  H'_j(w_n) &= 0 \quad (A.4)
\end{align*}
\]

As in section 2, \( h^*_n \) denotes the optimal amount of labor supply if the budget constraint is linear with slope \( w_{nj} \) and intercept \( \mu_{nj} \), \( j = 1, \ldots, m \):

\[
h^*_n = h^*(\alpha, \mu_{nj}, \zeta, q_n) + \alpha_2 v_n \quad (A.5)
\]

where

\[
h^*(\alpha, w, \mu, \zeta, q) = \alpha_1 + \alpha_2 \mu + \alpha_3 w + \frac{1}{2} \alpha_4 w^2 + q^\zeta \quad (A.6)
\]

Expressing \( \alpha \) and \( \zeta \) in terms of the original parameters gives:

\[
\begin{align*}
  \alpha_1 &= \delta + \beta \theta_0 \quad (A.7) \\
  \alpha_2 &= \beta \quad (A.8) \\
  \alpha_3 &= \gamma + \beta \delta \quad (A.9) \\
  \alpha_4 &= \beta \gamma \quad (A.10) \\
  \zeta &= \omega \beta \quad (A.11)
\end{align*}
\]

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

\[
h^*_{nj} = h^*(\alpha, \mu_{nj}, \zeta, q_n) \quad (A.12)
\]

For reasons explained in section 2, the unobserved taste parameter \( v_n \) is assumed to be distributed according to a negative gamma-distribution (i.e., \(-v_n\) has a gamma distribution) with parameters \( \gamma_1 \) and \( \gamma_2 \), i.i.d. over individuals. The probability density function of \( v_n \) is

\[
g(v_n, \gamma_1, \gamma_2) = \frac{1}{\Gamma(\gamma_1)\gamma_2^\gamma} (v_n)^{\gamma_1 - 1} \exp\left(\frac{v_n}{\gamma_2}\right), \gamma_1 > 0, \gamma_2 > 0, -\infty < v_n < 0 \quad (A.13)
\]

As pointed out in section 2, the distribution of the measurement errors is assumed to be normal with mean zero and variance \( \sigma_n^2 \).

\[
\phi(\epsilon_n, \sigma_n^2) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(\frac{-1}{2\sigma_n^2} \epsilon_n^2\right), -\infty < \epsilon_n < \infty \quad (A.14)
\]

Using the assumptions in (2.19),(2.20) and (2.21) the density function of the gross wage rate reads

\[
\begin{align*}
  \kappa(\omega_n, \eta, \tau^2) &= \frac{1}{\sqrt{2\pi\tau^2}} \frac{\bar{\omega}}{\bar{\omega} - \omega_n} \frac{1}{\omega_n} \exp\left\{ -\frac{1}{2\tau^2} \left[ \log\left(\frac{\omega_n}{\bar{\omega} - \omega_n}\right) - m_n \right]^2 \right\} \quad (A.15) \\
  m_n &= -\log\left(\frac{\bar{\omega}}{\exp(\eta n)} - 1\right), 0 < \omega_n < \bar{\omega} \quad (A.16)
\end{align*}
\]
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. In the estimation on the real data, such correlation between wages and measurement errors is allowed for. In the Monte Carlo experiments this correlation has been set equal to zero for the sake of simplicity.

The likelihood contribution of an individual will be formulated now. The algorithm (2.22) is used to determine optimal labor supply. Subsequently the participation rules described in (2.13) and (2.14) can be used to determine (simulated) observed labour supply. First note that the values for \( h^j_n \) in (2.22) all depend on the random preference parameter \( v_n \), so given everything else, \( v_n \) determines in which segment of the budget constraint labor 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:

\[
A_j = \{ v_n | H_{n,j-1} < h^j_n \leq H_{nj} \} \quad j = 0, \ldots, m
\]

\[
B_0 = \{ v_n | h^0_n \leq 0 \}
\]

\[
B_j = \{ v_n | h^j_n > H_{nj} > h^{j+1}_n \} \quad j = 1, \ldots, m - 1
\]

\[
B_m = \{ v_n | h^m_n \geq T \}
\]

\[
H_{n,-1} = 0, H_{nm} = T
\]

\[
Q(h_{n,NC}, h_{nj}) = \{ v_n | U(h_{n,NC}, y_0(h_{n,NC})) < U(h_{nj}, y_j(h_{nj})) \}
\]

\[
Q^*(h_{n,NC}, h_{nj}) = \{ v_n | U(h_{n,NC}, y_0(h_{n,NC})) > U(h_{nj}, y_j(h_{nj})) \}
\]

with \( h_{n,NC} \) defined in (2.22) and \( y_j(h) = w_j h + \mu_j \)

\[
R_j(h_{n,NC}, h^j_n) = A_j \cap Q^*(h_{n,NC}, h^j_n) \quad j = 1, \ldots, m
\]

\[
S_j(h_{n,NC}) = B_j \cap Q^*(h_{n,NC}, h^j_n) \quad j = 1, \ldots, m
\]

\[
Z_{01} = B_0 \cap \left( \bigcup_{j=1}^{m} R_j(0, h^j_n) \right) \cup \left( \bigcup_{j=1}^{m} S_j(0) \right)
\]

\[
Z_{02} = A_0 \cap \left( \bigcup_{j=1}^{m} R_j(h^0_n, h^n_n) \right) \cup \left( \bigcup_{j=1}^{m} S_j(h^0_n) \right)
\]

\[
Z_{j1} = B_j \cap Q(h_{n,NC}, H_{nj}) \quad j = 1, \ldots, m
\]

\[
Z_{j2} = A_j \cap Q(h_{n,NC}, h^j_n) \quad j = 1, \ldots, m
\]

\[
Z^* = \left( \bigcup_{j=1}^{m} Z_{j1} \right) \cup \left( \bigcup_{j=0}^{m} Z_{j2} \right)
\]

Z_{01} is the set of \( v_n \) for which optimal labor 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 labor supply is equal to the j-th kink, \( j = 1, \ldots, m \), \( Z_{j2} \) is the set for which it is optimal to be on the j-th segment after the nonconvexity kink, \( j = 1, \ldots, m \) and \( Z^* \) is the set for which optimal labor supply is positive.

We now determine the probability that observed labor supply is zero, conditional on the value of \( v_n \). According to (2.14) there are two possibilities for observed labor supply to be zero. The first happens when optimal labor supply is zero. Then observed labor supply is equal to zero, irrespective
of the value of measurement error. So if \( v_n \) is from the set for which optimal labor supply is zero, the probability that observed labor supply is zero, conditional on \( v_n \), is equal to one. The second possibility for observed labor supply to be zero occurs when optimal labor supply is positive but optimal labor supply plus measurement error is negative. Summarizing, the probability that observed labor supply is zero, conditional on \( v_n \) becomes:

\[
P(h_n = 0|v_n, w_n) = \begin{align*}
1 & \quad \text{if } v_n \in Z_0 \\
\Phi \left( -\frac{H_n}{\sigma_e} \right) & \quad \text{if } v_n \in Z_{j1}, j = 1, \ldots, m \\
\Phi \left( -\frac{h_i}{\sigma_e} \right) & \quad \text{if } v_n \in Z_{j2}, j = 0, \ldots, m
\end{align*}
\]

(A.18)

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

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

\[
\begin{align*}
\chi(h_n|v_n, w_n) &= \phi(h_n - H_{nj}, \sigma_e^2) \quad \text{if } v_n \in Z_{j1}, j = 1, \ldots, m \\
\chi(h_n|v_n, w_n) &= \phi(h_n - h_i, \sigma_e^2) \quad \text{if } v_n \in Z_{j2}, j = 0, \ldots, m
\end{align*}
\]

(A.19)

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

\[
P(h_n = 0|w_n) = \begin{align*}
\int_{Z^* \cup Z_0} P(h_n = 0|v, w_n)g(v, \gamma_1, \gamma_2)dv & \quad \text{if } h_n = 0 \\
\int_Z \chi(h_n|v, w_n)g(v, \gamma_1, \gamma_2)dv & \quad \text{if } h_n > 0
\end{align*}
\]

(A.20)

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

\[
P(h_n = 0|w_n) = \begin{align*}
\int_{Z_0} g(v, \gamma_1, \gamma_2)dv + \sum_{j=1}^m \int_{Z_{j1}} \Phi \left( -\frac{H_{nj}}{\sigma_e} \right) g(v, \gamma_1, \gamma_2)dv + \sum_{j=0}^m \int_{Z_{j2}} \Phi \left( -\frac{h_i}{\sigma_e} \right) g(v, \gamma_1, \gamma_2)dv & \quad \text{if } h_n = 0 \\
\sum_{j=1}^m \int_{Z_{j1}} \phi(h_n - H_{nj}, \sigma_e^2)g(v, \gamma_1, \gamma_2)dv + \sum_{j=0}^m \int_{Z_{j2}} \phi(h_n - h_i, \sigma_e^2)g(v, \gamma_1, \gamma_2)dv & \quad \text{if } h_n > 0
\end{align*}
\]

(A.21)

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

\[
\int_0^w P(h_n = 0|w)\kappa(w, \eta, \tau^2)dw
\]

(A.23)

The problem with the sets defined above 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 labor supply is.

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 numerical optimization of the objective function. These basic drawings can be transformed to drawings from the distribution of \( \epsilon_n \) through multiplying by \( \sigma_\epsilon \). Any change in the drawings of \( \epsilon_n \) is caused by a change in \( \sigma_\epsilon \).

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

\[
\tilde{w}_{nr} = \frac{\tilde{w}_n \exp(m_n + \tau \tilde{l}_{nr})}{1 + \exp(m_n + \tau \tilde{l}_{nr})}, r = 1, \ldots, R
\]  

(A.24)

The transformation is continuous in the parameters and therefore, keeping the basic drawings constant, a change in the drawings \( \tilde{w}_{nr} \) 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 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 \( \rho \):

\[
\Lambda(\rho, v_n) = \rho \exp\{\rho v_n\}, -\infty < v_n < 0, \rho > 0
\]  

(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.

\[
k(v_n, \gamma_1, \gamma_2, \rho) = \frac{g(v_n, \gamma_1, \gamma_2)}{\Lambda(\rho, v_n)} = \frac{1}{(\gamma_1)\gamma_2(\rho)^{\gamma_1}} (-v_n)^{\gamma_1 - 1} \exp \left\{ \left( \frac{1}{\gamma_2} - \rho \right) v_n \right\}
\]  

(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 \( \rho \) in such a way that the variance will be finite and second, the choice of \( \rho \) 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 we is draw 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:

\[
E[k(v, \gamma_1, \gamma_2, \rho)]^2 - 1 = \\
\int_{-\infty}^{0} \left[ \frac{g(v, \gamma_1, \gamma_2)}{\Lambda(v, \rho)} \right]^2 \Lambda(v, \rho) dv - 1 = \\
\int_{-\infty}^{0} k(v, \gamma_1, \gamma_2, \rho) g(v, \gamma_1, \gamma_2) dv - 1 = \\
\frac{\Gamma(2\gamma_1 - 1) \left( \frac{\gamma_2}{\gamma_1} \right)^{2\gamma_1 - 1}}{[\Gamma(\gamma_1)]^{2\gamma_2 / \gamma_1 \rho}} - 1
\]

(A.27)

in which

\[
\gamma_1 > \frac{1}{2} \quad \text{(A.28)} \\
\rho < \frac{2}{\gamma_2} \quad \text{(A.29)}
\]

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 \( \rho \) 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

\[
\rho = \frac{1}{\gamma_1 \gamma_2} \quad \text{(A.30)}
\]

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

In summary, 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 \( \rho \). 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 \( \gamma_1 \gamma_2 \):

\[
v_{nr}^* = \gamma_1 \gamma_2 \tilde{v}_{nr}^*
\]

(A.31)

Note that this is a continuous transformation in \( \gamma_1 \) and \( \gamma_2 \). These are the final drawings which will be used in the simulation of the labor supply.

Having described the generation of the required random numbers, we now turn to the simulation of the probability of participation. Using the drawings \( (\epsilon_{nr}^*, v_{nr}^*, w_{nr}^*) \) the optimal labor supply \( \hat{h}_{nr} \) and the observed labor supply \( h_{nr} \) can be simulated according to the algorithm (2.22) and the participation rules (2.13) and (2.14). Then the participation probability can be simulated by a frequency simulator like in (2.23)-(2.25) where (2.23) has to be weighted. The frequency simulator becomes:

\[
f_{nr} = k(v_{nr}^*, \gamma_1, \gamma_2, \rho) \quad \text{if} \ h_{nr} > 0 \\
f_{nr} = 0 \quad \text{otherwise}
\]

(A.32)

(A.33)

\[
F_{nR} = \frac{1}{R} \sum_{r=1}^{R} f_{nr}
\]

(A.34)

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 of the participation probability at the parameter vector \( \theta \). Then the derivative with respect to the \( k \)-th component of \( \theta \) is simulated by a difference interval of frequency simulators:

\[
\tilde{m}_{nk}(\theta, \epsilon_{R}^*, v_{R}^*, w_{R}^*) = \frac{F_{nR}(\theta + \delta \epsilon_k) - F_{nR}(\theta)}{\delta}
\]

(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 $\theta$, (A.35) is an unbiased simulator of the difference interval of the participation probability. Because $F_{nR}(\theta)$ is discontinuous in the parameter vector $\theta$ we have to choose $\delta$ 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 $\delta$. 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 2000 drawings, and 800 in the Monte Carlo experiment. In the application we have chosen a difference interval of 0.0001.

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 a gamma distribution. The derivative of the continuous part of the log-likelihood function is

$$
\frac{\partial \ln l(h_n|w_n)}{\partial \theta} = \frac{\partial l(h_n|w_n)}{\partial \theta}/l(h_n|w_n)
$$

(A.36)

Numerator and denominator in (A.36) can be simulated separately by drawing random numbers from the density $g(v, \gamma_1, \gamma_2)$. We illustrate this in detail for the denominator. Draw random numbers $v_{nr}^*$ from the density $g(v, \gamma_1, \gamma_2)$ restricted to the region $Z^*$ for which optimal labor supply is positive, defined in (A.17).

$$
v_{nr}^* \sim g(v, \gamma_1, \gamma_2) \quad v_{nr}^* \in Z^*
$$

(A.37)

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

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

(A.38)

or, writing this out:

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

$$
P(v \in Z^*) \frac{1}{R} \sum_{r=1}^{R} \left\{ \sum_{j=0}^{m} I(v_{nr}^* \in Z_{j2})\phi(h_n - h_{nj}, \sigma_n^2) + \sum_{j=1}^{m} I(v_{nr}^* \in Z_{j1})\phi(h_n - H_{nj}, \sigma_n^2) \right\}
$$

(A.39)

in which $h_{nj}$ is computed on the basis of $v_{nr}^*$ and $I(.)$ the indicator function. A simulator for numerator in (A.36) can be obtained in a similar way. The separate simulation of numerator and denominator introduces denominator bias in the continuous part of the score.

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 $\hat{Z}^*$ which contains the original region, i.e., $Z^* \subset \hat{Z}^*$, and which approximates the original region as close as possible. Consider the region

$$
\hat{Z}^* = \{v| - \infty < v < q(\alpha, \zeta, w_n, \mu_n)\}
$$

$$
q(\alpha, \zeta, w_n, \mu_n) =
$$

$$
-\frac{h_{n1}^*}{\alpha_2} if -\frac{h_{n1}^*}{\alpha_2} < 0
$$

$$
0 if -\frac{h_{n1}^*}{\alpha_2} > 0
$$

(A.40)

$-\frac{h_{n1}^*}{\alpha_2}$ is the value of $v$ for which $h_{n1}^*$ is equal to zero. The region $Z^*$ of positive optimal labor 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 $\rho$:

$$
v_{nr}^* \sim \rho \exp\{\rho v_{nr}^*\}, -\infty < v_{nr}^* < 0
$$

(A.41)
Define $\nu^*_{nr}$ as

$$
\nu^*_{nr} = \nu^*_{nr} + q(\alpha, \zeta, w_n, \mu_n)
$$

(A.42)

As a consequence, $\nu^*_{nr}$ is a truncated negative exponential variable:

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

(A.43)

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

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

(A.44)
References


