The joint estimation of a non-linear labour supply function and a wage equation using simulated response probabilities
Bloemen, H.G.; Kapteyn, A.J.

Publication date: 1993

Citation for published version (APA):
The Joint Estimation of a Non-Linear Labour Supply Function and a Wage Equation Using Simulated Response Probabilities

by

Hans G. Bloemen and Arie Kapteyn

Reprinted from Annales d'Économie et de Statistique, No. 29, 1993

Reprint Series no. 136
CENTER FOR ECONOMIC RESEARCH

Board
Harry Barkema
Helmut Bester
Eric van Damme, chairman
Frank van der Duyn Schouten
Jeffrey James

Management
Eric van Damme (graduate education)
Arie Kapteyn (scientific director)
Marie-Louise Kemperman (administration)

Scientific Council
Anton Barten
Eduard Bomhoff
Willem Buiter
Jacques Drèze
Theo van de Klundert
Simon Kuipers
Jean-Jacques Lafont
Merton Miller
Stephen Nickell
Pieter Ruys
Jacques Sijben

Université Catholique de Louvain
Erasmus University Rotterdam
Yale University
Université Catholique de Louvain
Tilburg University
Groningen University
Université des Sciences Sociales de Toulouse
University of Chicago
University of Oxford
Tilburg University
Tilburg University

Residential Fellows
Lans Bovenberg
Werner Güth
Jan Magnus
Shigeo Muto
Theodore To
Karl Wärneryd
Karl-Erik Wärneryd

CentER, Erasmus University Rotterdam
University of Frankfurt
CentER, LSE
Tohoku University
University of Pittsburgh
Stockholm School of Economics
Stockholm School of Economics

Research Coordinators
Eric van Damme
Frank van der Duyn Schouten
Harry Huizinga
Arie Kapteyn

Address : Warandelaan 2, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
Phone : +31 13 663050
Telex : 52426 kub nl
Telefax : +31 13 663066
E-mail : "center@htikub5.bitnet"

ISSN 0924-7874
The Joint Estimation of a Non-Linear Labour Supply Function and a Wage Equation Using Simulated Response Probabilities

by

Hans G. Bloemen and Arie Kapteyn

Reprinted from Annales d’Économie et de Statistique, No. 29, 1993

Reprint Series no. 136
The Joint Estimation of a Non-Linear Labour Supply Function and a Wage Equation Using Simulated Response Probabilities

Hans G. BLOEMEN, Arie KAPTEYN

ABSTRACT. — When applying maximum likelihood estimation in jointly estimating a labour supply function and a wage equation, it may be practically impossible, both analytically and numerically, to calculate the required response probabilities, especially if the model is non-linear. As an alternative, we consider various simulation estimators. In both Monte Carlo experiments and empirical applications the methods are compared to each other and to ML. The methods are computationally feasible and perform well.

L’estimation jointe d’une fonction non-linéaire d’offre de travail et d’une équation de salaire utilisant des probabilités de réponses simulées

RÉSUMÉ. — Lorsqu’on applique une méthode de vraisemblance pour estimer conjointement une fonction d’offre de travail et une équation de salaire, en pratique, il peut être impossible de calculer les probabilités de réponses requises, à la fois analytiquement et numériquement, plus particulièrement si le modèle est non-linéaire. Pour traiter ce problème nous utilisons différents estimateurs construits sur des méthodes de simulation. Nous comparons les résultats des méthodes de Monte-Carlo et les applications empiriques entre elles et avec les estimations du maximum de vraisemblance. Ces méthodes donnent des résultats satisfaisants et leur coût de programmation reste modéré.

* Hans G. BLOEMEN, A. KAPTEYN: center for economic research, Tilburg University. The authors thank the Organisatie voor Strategisch Arbeidsmarktonderzoek (OSA) for kindly providing the data. Thanks are due to Arthur van Soest for his help and to Peter Kooreman for his useful comments.
1 Introduction

This paper investigates the applicability of estimation methods for labour supply models which make use of simulators for the response probabilities. The methods of estimation usually applied, like maximum likelihood and the method of moments, make use of the probabilities of individuals participating in the labour force. If the labour supply model is non-linear and if one wants to incorporate the tax and social security system, thereby assuming that the budget constraints of the individuals are non-convex, the calculation of the participation probabilities may be impossible and there is possibly no other way to estimate the model than making use of simulated moments types of estimators.

McFaddens [1989] presents a method of simulated moments estimator for the multinomial response model. The attractiveness of the method is that the number of replications which is used to simulate the response probabilities can be kept fixed to any positive integer without destroying the consistency property of the estimator. In a short time an extensive literature has blossomed in which this approach has been extended and refined. The emphasis has been on computational accuracy and speed in the evaluation of multi-dimensional probabilities, often under normality or closely related assumptions and linearity. See Hajiassililou [1992] for an overview.

The method of simulated moments estimator, however, is in its simplest form only suitable for discrete response data, whereas in labour supply models the data are usually of a mixed discrete-continuous nature, giving information on whether or not individuals are working and if so, how many hours. Furthermore, any realistic utility consistent model will entail non-linearities and non-normality, so that the various refinements mentioned will not be applicable. Therefore, we want to set up estimation by simulation methods for the mixed discrete-continuous type of model that one typically finds in labour supply analysis.

Different routes can be followed. The most straightforward way is to replace the response probabilities in the likelihood function by simulators, thereby simulating the likelihood function. It can be inferred from Gouriéroux and Montfort [1989] that this method of simulated maximum likelihood (SML) is not consistent for an arbitrary, fixed number of replications to simulate the response probabilities. Lerman and Manski [1981] show that a similar method for the multinomial response model may require huge numbers of replications. To circumvent this problem an alternative is to use the method of simulated scores (MSS) which is based on the simulation of the vector of scores of the log-likelihood function. As in McFadden [1989], point of departure is the property of the likelihood function that under weak regularity conditions the expectation of the score vector equals zero at the true parameter value. This score vector will be replaced by a simulated score vector. An estimator can be obtained by
minimizing the length of the simulated score. The score vector will be simulated in such a way that the property of having a zero expectation at the true parameter value carries over to the simulated score vector. There is no unique way to achieve this and therefore we will propose and compare three different methods of estimation. The method of simulated scores is also used by Hajivassiliou [1989].

The methods of estimation will be applied to the joint estimation of a labour supply function, non-linear in the wage rate, and a wage equation. In this application, we assume a linear budget constraint. This is a rather simple model and in fact it can be estimated by maximum likelihood, using numerical integration. The main purpose of the application is to gain insight in the practical properties of the MSS methods. Hence we have chosen a model simple enough so that ML is feasible and we can compare the performance of the MSS estimators with ML. In a companion paper (Bloemen and Kapteyn [1992]) one of the MSS estimators is applied to a much more complicated model with random preferences and non-convex budget constraints. In that model ML is not feasible. We will present Monte Carlo results as well as real data estimates. In the Monte Carlo study different MSS estimators are compared with each other as well as with ML and SML with a limited number of replications.

The order of presentation is as follows: In the next section we set out the basic model where labour supply is a (possibly nonlinear) function of the wage rate and nonlabour income. Errors in the wage equation are additive and their nature remains unspecified. In section 3 different simulators for the score vector are proposed, each of them generating an alternative method of estimation. Also, attention will be paid to the statistical properties of the methods. In Section 4 the properties of these estimators and ML are first investigated by means of some Monte Carlo experiments. Next, the estimators are applied to the analysis of labour supply of Dutch females.

The general finding is that the MSS estimators perform quite well, though of course slightly below ML, whereas simulated ML may perform poorly. We conclude that the MSS estimators proposed present viable routes for the estimation of utility consistent labour supply models.

2 The Basic Model

Our point of departure is a two-equation model consisting of a labour supply equation and a log-wage equation.

\[ h_n^* = h(w_n, \mu_n; \beta) + \varepsilon_n \]
\[ \log(w_n) = w(x_n, \eta) + u_n \]
\( h_n^* \) is observed and equals the number of hours worked by individual \( n \) if the \( n \)-th individual is working; \( h_n^* \) is unobserved if the \( n \)-th individual is non-working.

\[
\begin{align*}
(3) & \quad h_n = h_n^* \quad \text{if} \quad h_n^* > 0 \\
(4) & \quad h_n = 0 \quad \text{if} \quad h_n^* \leq 0
\end{align*}
\]

where \( h_n \) is the actual number of hours worked by individual \( n \), \( w_n \) is the after-tax real wage rate which will be unobserved for a non-working individual, \( \mu_n \) is non-labour income, \( x_n \) is a vector of observable characteristics of individual \( n \), \( \beta \) is a parameter vector with dimension \( l \), \( \eta \) is a vector of parameters with dimension \( q \), \( \epsilon_n \) and \( u_n \) are random disturbances with expectation 0 and covariance matrix

\[
\Sigma = \begin{pmatrix}
\sigma^2_\epsilon & \sigma_{\epsilon u} \\
\sigma_{\epsilon u} & \sigma^2_u
\end{pmatrix}
\]

and joint probability density function \( f(\epsilon_n, u_n) \), independent across observations. For ease of notation we introduce the dummy variable \( d_n \) with

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

We start by deriving the joint probability density function of \( h_n \) and \( w_n \) given \( x_n \) and \( \mu_n \). First, an expression for the joint density of \( h_n^* \) and \( w_n \) has to be found. Using the 1-1 transformations \( \epsilon_n = h_n^* - h(w_n, \mu_n; \beta) \) and \( u_n = \log(w_n) - w(x_n; \eta) \), we can employ the joint density of \( \epsilon_n \) and \( u_n \) to get the density function \( g^*(h_n^*, w_n) \) of \( h_n^* \) and \( w_n \). The Jacobian of the transformation is

\[
\begin{vmatrix}
\frac{\partial \epsilon_n}{\partial h_n^*} & \frac{\partial u_n}{\partial h_n^*} \\
\frac{\partial \epsilon_n}{\partial w_n} & \frac{\partial u_n}{\partial w_n}
\end{vmatrix} = \begin{vmatrix}
1 & 0 \\
-\frac{\partial h}{\partial h_n^*} & 1
\end{vmatrix} \frac{1}{w_n}
\]

So

\[
(7) \quad g^*(h_n^*, w_n) = f(h_n^* - h(w_n, \mu_n; \beta), \log(w_n) - w(x_n; \eta)) \frac{1}{w_n}
\]

\[-\infty < h_n^* < \infty \]

\[0 < w_n < \infty \]

From this we can derive the mixed discrete-continuous probability density function of \( h_n \) and \( w_n \), \( g(h_n, w_n | x_n, \mu_n, \theta) \) where \( \theta \) contains the parameters of \( \beta, \eta \) and the upper triangular or, equivalently, the lower triangular elements of \( \Sigma \).

\[
g(h_n, w_n | x_n, \mu_n, \theta) =\begin{cases} 
P(h_n^* \leq 0 | x_n, \mu_n, \theta) & \text{if} \quad h_n = 0 \\
g^*(h_n^*, w_n | x_n, \mu_n, \theta) & \text{if} \quad h_n > 0, \quad 0 < w_n < \infty \end{cases}
\]
where

\[ P(h_n^* \leq 0 | x_n, \mu_n, \theta) = \int_0^{\infty} \int_{-\infty}^{0} g^*(h, w | x_n, \mu_n, \theta) \, dh \, dw \]

For ease of notation this probability will be denoted by \( P_n(\theta) \) or by \( P_\theta \). The wage \( w \) is integrated out because for non-working individuals we have no observations on the wage rate. We shall denote the probability of working \( 1 - P_n(\theta) \) by \( P_n(\theta) \) or simply by \( P_n \). We assume that our sample is ordered in such a way that the observations 1 to \( N \) refer to non-working individuals and the observations \( N + 1 \) to \( N \) are working individuals.

We now formulate the log-likelihood function of the model.

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

This is differentiated with respect to \( \theta \) to derive the first order conditions for a maximum.

\[ \frac{\partial L(\theta)}{\partial \theta} = \sum_{n=1}^{N} \frac{\partial \ln P_n(\theta)}{\partial \theta} + \sum_{n=N+1}^{N} \frac{\partial \ln g^*(h_n, w_n | x_n, \mu_n, \theta)}{\partial \theta} \]

(11) \[ \frac{\partial L(\hat{\theta}_{ML})}{\partial \theta} = 0 \]

where \( \hat{\theta}_{ML} \) is the maximum likelihood estimator of \( \theta \).

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

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

where \( d_n \) is the dummy variable introduced above.

Let \( \theta_0 \) be the true parameter value. It is easy to show that if the supports of \( h_n \) and \( w_n \) do not depend on \( \theta \),

\[ E\left( \frac{\partial L(\theta_0)}{\partial \theta} \right) = 0 \]

which is the result of the fact that the expectation of the derivative of the log-density function with respect to \( \theta \) at the true parameter value equals zero. In fact, the first order derivative of the log-likelihood function divided by the sample size can be looked upon as a moment estimator of

\[ E\left( \frac{\partial \ln g(h, w | x_n, \mu, \theta)}{\partial \theta} \right) \]

evaluated at the parameter value \( \theta \).
A procedure which is often followed in estimating this type of model is a two-step procedure. First, the wage equation is estimated using data on working individuals. The resulting estimates of the parameter vector \( \eta \) and the characteristics \( x \) of the non-working individuals are used to construct a proxy for the wage variables of the non-working individuals. Second, this constructed data-set is used to estimate the labour supply function, using Tobit-like methods. This method will in general yield inconsistent estimates, particularly for non-linear labour supply functions. A correct procedure would be to estimate the model simultaneously. A drawback of simultaneous estimation of the model is the difficulty in calculating the response probabilities analytically whenever the model is non-linear, whereas numerical approximation can be expensive and time-consuming. Our purpose is to develop estimation methods for models of the type (1)-(2) that allow for simultaneous estimation also if the function \( h \) is quite complicated (as for instance in the case, where \( h \) represents the outcome of utility maximization under a non-linear and non-convex budget constraint). To this end we make use of simulators for the response probabilities like the ones proposed by McFadden [1989]. However, unlike his paper, we shall not restrict ourselves to the discrete response model. We will employ a discrete-continuous type of estimation method. The property (13) will be made use of to develop a method of simulated moments type of estimation. We want to replace the response probabilities in (12) by simulators and we want to do that in such a way that property (13) carries over to the simulated score vector. Then an estimator can be found by minimizing the length of simulated score vector.

3 Estimation

Three ways of simulating the score of the log-likelihood function are considered. The first method replaces the discrete part of the score by an expression with an instrument matrix and a simulator for the response probability of non-working individuals. The disadvantage of this method is that the estimator will depend on the choice of the matrix of instruments. This is due to the fact that the expectation of the simulated score, evaluated at the true parameter value \( \theta_0 \), does not equal zero, unless a specific form for the matrix of instruments is chosen which makes use of a consistent estimator for \( \theta_0 \). So in order to get a consistent estimator, we need a consistent estimator obtained from a different estimation procedure. Therefore, the first method is only useful to increase efficiency of the first round estimator obtained by one of the next two methods. The second method of estimation also uses a matrix of instruments. The estimator will be consistent, irrespective of the choice of the instruments. To
simulate the score, simulators of the response probabilities and their derivatives are needed for each individual, both non-working and working. A second estimation round can be performed to increase the efficiency of the estimators, using an updated version of the matrix of instruments. This method directly extends McFadden's [1989] estimation method for the discrete response model, by adding a continuous component to his objective function. The third method does not rely on a matrix of instruments. Only simulators of the derivatives of the response probabilities are required.

3.1. Method 1

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

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

where

\[
Z_n := \frac{\partial P_n/\partial \theta}{P_n (1 - P_n)}
\]

Now the vector \( Z_n \) is replaced by an arbitrary vector of instruments \( Z_{n*} \), which does not depend on \( \theta \). The resulting expression is

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

where the superscript 1 refers to the number of the method. Calculating the expectation of expression (17), conditional on \( Z_{n*} \), evaluated at the true parameter value \( \theta_0 \) yields

\[
E\left( \frac{\partial L^1}{\partial \theta} | Z_{n*} \right) = \sum_{n=1}^{N} \left[ P_n Z_n (1 - P_n) + \frac{\partial P_n}{\partial \theta} \right]
\]

(with \( P_n = 1 - P_n \)) which in general doesn’t equal zero. However, if we construct \( Z_{n*} \) in such a way that

\[
\lim_{P_n \to 0} \frac{\partial P_n/\partial \theta}{P_n (1 - P_n)} \quad \text{at} \quad \theta_0
\]

we have that at the true parameter value \( \theta_0 \) the simulated score has asymptotic mean equal to zero.

The response probability \( P_n(\theta) \) in expression (17) is replaced by a frequency simulator or by a so called smooth unbiased simulator \( k_n(\theta, v_R^n) \) where \( v_R^n \) is a vector of R drawings from a distribution. The simulator is
unbiased if it has the property

\[ E(k_n(\theta, v^*_R)) = P_n(\theta) \]

A smooth unbiased simulator can be constructed in the same way as in Monte Carlo importance sampling, see e.g. **Hammersley** and **Handscmb** [1979]. Take a density function \( \gamma(h, w) \) with support coinciding with the bounds of the integrals in the expression for \( P_n \). The response probability can be rewritten as:

\[ P_n(\theta) = \int_0^{\infty} \int_{-\infty}^{\infty} \tau(h, w | x_n, \mu_n, \theta) \gamma(h, w) \, dh \, dw \]

with

\[ \tau(h, w | x_n, \mu_n, \theta) = \frac{g^*(h, w | x_n, \mu_n, \theta)}{\gamma(h, w)} \]

For every \( n \), \( R \) random numbers \( h_{(r,n)}, w_{(r,n)} \) are drawn from the density function \( \gamma(h, w) \), independently across observations and not depending on the parameter vector \( \theta \). By averaging over the drawings a simulator is obtained:

\[ k_n(\theta, v^*_R) = \frac{1}{R} \sum_{r=1}^{R} \tau(h_{(r,n)}, w_{(r,n)} | x_n, \mu_n, \theta) \]

where \( v^*_R \) consists of the drawings \( h_{(r,n)} \) and \( w_{(r,n)} \). The function \( \tau(\ldots | \ldots, \ldots) \) is the so called weight function which corrects for the fact that we are drawing random numbers from the distribution with density function \( \gamma(\ldots) \) instead of the true distribution with density function \( g^*(\ldots | \ldots, \ldots) \). If \( \gamma(\ldots) \) and \( g^*(\ldots | \ldots, \ldots) \) coincide the weight function is identically equal to 1. In our application, described in the next section, we will actually use a slightly different way of simulating the response probabilities by exploiting the normality assumptions and the assumed linearity of the log-wage equation. Therefore, we don't have to choose a density function \( \gamma(h, w) \). However, in more complicated applications, as in **Bloemen** and **Kapteyn** [1992], weight functions are necessary.

As indicated by **McFadden** [1989], a simulator for the derivatives of \( P_n \) can be constructed in a similar way. An unbiased simulator \( m_n(\theta, v^*_R) \) of the derivatives of \( P_n \) is

\[ m_n(\theta, v^*_R) = \frac{1}{R} \sum_{r=1}^{R} \frac{\partial \tau(h_{(r,n)}, w_{(r,n)} | x_n, \mu_n, \theta)}{\partial \theta} \]

For the simulation of (17) we only need \( k_n \). The simulated score is:

\[ K_R^I(\theta) = \sum_{n=1}^{N} \left[ d_n Z_n(1 - k_n(\theta, v^*_R)) + (1 - d_n) \frac{\partial \ln g^*(h_n, w_n | x_n, \mu_n, \theta)}{\partial \theta} \right] \]
The estimation procedure now becomes: Minimize the length of the simulated score vector:

\[ \min_{\theta} s_N(\theta) \]

where

\[ s_N(\theta) = K^1_R(\theta)' K_R^1(\theta) \]

where the matrix of instruments has to be based on the following formula, evaluated in the true parameter point (or at least at a consistent estimate of it):

\[ Z_n = \frac{m_n(\theta, v_{RZ}^*)}{k_n(\theta, v_{RZ}^*)(1 - k_n(\theta, v_{RZ}^*))} \]

where \( m_n(\theta, v_{RZ}) \) is an unbiased simulator of the derivatives of \( P_n(\theta) \), and \( v_{RZ}^* \) are drawn independently of the \( v_R^* \) which are used in minimizing \( s_N(\theta) \). The number of drawings to simulate the vector of instruments is denoted by \( R_Z \) to indicate that it is not necessarily equal to \( R \), the number of drawings used to construct \( k_n(\theta, v_R^*) \) in (25). The reason for presenting this method of estimation is that the asymptotic variance of this method is lower than the asymptotic variance of the methods presented in the next two subsections, which will be explained in section 3.4. Therefore, a two-step procedure could be followed: First, obtain a consistent estimate by applying one of the other estimation methods and second, use the estimates to construct the vector of instruments in (28) and apply method I to increase the efficiency of the estimates.

We are interested in the error we make by replacing the score vector by a simulator. Therefore, the simulated score in (25) is rewritten as the sum of the true score in (15) and a simulation residual.

\[ K^1_R(\theta) = \frac{\partial L}{\partial \theta} + \text{RES}_1 \]

where

\[ \text{RES}_1 = \sum_{n=1}^{N} d_n Z_n (P_n - k_n) + \sum_{n=1}^{N} d_n (Z_n - Z_n)(1 - k_n) \]

For ease of notation the arguments are omitted. The first term of (30) can be rewritten as

\[ \frac{1}{R} \sum_{n=1}^{N} \sum_{r=1}^{R} d_n Z_n (P_n - k_{nr}) \]

where \( k_{nr} \) is the \( r \)-th term of \( k_n \). By increasing the number of drawings \( R \) this term will tend to zero. In the second term, the vector \( Z_n \) appears which is a non-linear function of the simulators. Because the two factors in this term are independent by construction, we concentrate on \( Z_n - Z_n \). If
Ze is constructed in a proper way, i.e. by using a consistent estimate for 0, then

\[ p \lim \limits_{R \to \infty} (Z_n - Z_0) = 0 \]

so by taking the number of drawings to construct the vector of instruments large enough the second term also can be made arbitrarily small. The variance of the simulation residual determines the loss of efficiency caused by applying the simulated score instead of the true score vector. In the next subsection we discuss how the variance of the simulation residual can be influenced by the number of drawings R.

Since the derivation of the asymptotic distribution is equivalent for the three estimators we will treat the asymptotic properties of the estimators at the end of this section.

### 3.2. Method 2

In order to obtain the second method of estimation we rewrite the score of the log-likelihood function as

\[ \frac{dL}{d\theta} = \sum_{n=1}^{N} \left\{ Z_n(d_n - P_n) + (1 - d_n) \left[ \frac{\partial \ln g^*(h_n, w_n | x_n, \mu_n, \theta)}{\partial \theta} - \frac{\partial \ln P_n}{\partial \theta} \right] \right\} \]

with \( Z_n \) defined as above. 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 \( \theta \), the expectation of the resulting expression, conditional on \( \tilde{Z}_n \), equals zero at the true parameter value \( \theta_0 \).

To simulate \( P_n(\theta) \) we use the simulator \( k_n(\theta, v_R^*) \) defined above. The problem is how to simulate \( \frac{\partial \ln P_n}{\partial \theta} \). To see why this is a problem we rewrite this expression as

\[ \frac{\partial \ln P_n}{\partial \theta} = \frac{1}{P_n} \frac{\partial P_n}{\partial \theta} \]

It is not difficult to construct unbiased simulators of \( P_n \) and \( \frac{\partial P_n}{\partial \theta} \), as we showed before. However, if we use their simulators \( \tilde{k}_n(\theta, v_R^*) \) and \( \tilde{m}_n(\theta, v_R^*) \) to simulate \( \frac{\partial \ln P_n}{\partial \theta} \), we don’t get an unbiased simulator.
So the expectation of the simulated score evaluated at \( \theta_0 \) won’t equal zero. It is not clear how to get an unbiased simulator of \( \frac{\partial \ln P_n}{\partial \theta} \). In order to solve this problem, we will simulate \( (1 - d_n) \frac{\partial \ln P_n}{\partial \theta} \) instead of \( \frac{\partial \ln P_n}{\partial \theta} \).

\[
E \left[ (1 - d_n) \frac{\partial \ln P_n}{\partial \theta} \right] = \frac{\partial P_n}{\partial \theta}
\]

We now replace \( (1 - d_n) \frac{\partial \ln P_n}{\partial \theta} \) by \( \frac{\partial P_n}{\partial \theta} \). As a result, the original score vector is replaced by

\[
\frac{\partial \ln^2 \theta}{\partial \theta} = \sum_{n=1}^{N} \left[ Z_n (d_n - P_n) + (1 - d_n) \frac{\partial \ln g^*(h_n, w_n \mid x_n, \mu_n, \theta)}{\partial \theta} - \frac{\partial P_n}{\partial \theta} \right]
\]

Inserting the simulators for the response probabilities and their derivatives in this expression gives the simulated score:

\[
K_R^2(\theta) = \sum_{n=1}^{N} \left[ Z_n (d_n - k_n(\theta, v_n^*)) + (1 - d_n) \frac{\partial \ln g^*(h_n, w_n \mid x_n, \mu_n, \theta)}{\partial \theta} - m_n(\theta, v_n^*) \right]
\]

The estimation procedure becomes: Choose instrument vectors \( Z_n \) and minimize the length of the simulated score.

\[
\min_\theta K_R^2(\theta)'
\]

We can increase the efficiency of the estimator by rerunning the procedure using the updated vectors of instruments, as described in the preceding subsection.

Again we are interested in the simulation residual. First, (33) with \( Z_n \) replaced by \( Z_n^* \) is compared with (38). Then the following residual is obtained:

\[
\frac{1}{R} \sum_{n=1}^{N} \sum_{r=1}^{R} \left[ Z_n (P_n - k_{nr}) - \left\{ m_{nr} - (1 - d_n) \frac{\partial \ln P_n}{\partial \theta} \right\} \right]
\]

This is the simulation residual corresponding to the comparison of the moments estimator (\( Z_n \) replaced by \( Z_n^* \)) with the simulated moments estimator. The dummy variable can be rewritten as

\[
d_n = P_n + v_n
\]

with \( E(v_n) = 0 \) and \( \text{Var}(v_n) = P_n (1 - P_n) \). 

NON-LINEAR LABOUR SUPPLY FUNCTION AND WAGE EQUATION 185
Inserting this in the residual gives:

\[
\frac{1}{R} \sum_{n=1}^{N} \sum_{r=1}^{R} \left[ Z_n(P_n - k_{nr}) - \left\{ \frac{\partial P_n}{\partial \theta} \right\} \right] - \sum_{n=1}^{N} v_n \frac{\partial \ln P_n}{\partial \theta}
\]  

The variance of the first term of (42) can be reduced by increasing the number of drawings R. Suppose that

\[
\text{Var} \left[ Z_n(P_n - k_{nr}) - \left\{ \frac{\partial P_n}{\partial \theta} \right\} \right] = \Xi_n,
\]

conditional on the instruments \( Z_n \). This variance does not depend on \( r \) because the drawings are i.i.d. Then the variance of the first term is

\[
\frac{1}{R} \sum_{n=1}^{N} \Xi_n.
\]

With fixed \( N \), increasing \( R \) to infinity results in reducing this variance to zero. The second term is the error which is caused by the fact that \((1 - d_n) \frac{\partial \ln P_n}{\partial \theta}\) is simulated by a simulator for \( \frac{\partial P_n}{\partial \theta} \). The expectation of this term equals zero, whereas the variance equals

\[
P_n \frac{\partial \ln P_n}{\partial \theta} \frac{\partial \ln P_n}{\partial \theta'}.
\]

This term of the simulation residual cannot be influenced by the number of drawings. Therefore, this term leads to inefficiency, also for large \( R \).

To compare the efficiency of the method of simulated moments estimator to the maximum likelihood estimator, also the term involving the difference between \( Z_n \) and \( Z^*_n \) has to be taken into account. The simulation residual then becomes:

\[
\text{RES}_2 = \frac{1}{R} \sum_{n=1}^{N} \sum_{r=1}^{R} \left[ Z_n(P_n - k_{nr}) - \left\{ \frac{\partial P_n}{\partial \theta} \right\} \right] + \sum_{n=1}^{N} (Z_n - Z^*_n)(d_n - k_n) - \sum_{n=1}^{N} v_n \frac{\partial \ln P_n}{\partial \theta}
\]

Here the same observations can be made as for method 1.

### 3.3. Method 3

As opposed to the first two methods of estimation, we won't rewrite the score of the log-likelihood function in a form involving instrument vectors \( Z_n \). Instead, we immediately replace the discrete part of the score by a simulator. In analogy with the preceding method, we replace \( d_n \frac{\partial \ln P_n}{\partial \theta} \) by
\[ \frac{\partial P_n}{\partial \theta} \] This yields the expression

\[ \frac{\partial L^3}{\partial \theta} = \sum_{n=1}^{N} \left[ \frac{\partial P_n}{\partial \theta} + (1 - d_n) \frac{\partial g^*(h_n, w_n \mid x_n, \mu_n, \theta)}{\partial \theta} \right] \]

The derivative of the probability is replaced by its simulator \( m_n(\theta, v_R^*) \), which leads to the following expression for the simulated score:

\[ K_R^3(\theta) = \sum_{n=1}^{N} \left[ m_n(\theta, v_R^*) + (1 - d_n) \frac{\partial g^*(h_n, w_n \mid x_n, \mu_n, \theta)}{\partial \theta} \right] \]

The estimation procedure becomes

\[ \min_{\theta} K_R^3(\theta) \]

We compare this method of simulating the score with the method described in the preceding subsection. Because \( P_n = 1 - P_n \), we find that

\[ \frac{\partial P_n}{\partial \theta} = - \frac{\partial P_n}{\partial \theta} \]

The same relation holds for their simulators:

\[ m_n(\theta, v_R^*) = - \tilde{m}_n(\theta, v_R^*) \]

Inserting this in the simulated score, we obtain

\[ K_R^3 = \sum_{n=1}^{N} \left[ (1 - d_n) \frac{\partial g^*(h_n, w_n \mid x_n, \mu_n, \theta)}{\partial \theta} - \tilde{m}_n(\theta, v_R^*) \right] \]

This is exactly the second component of the simulated score of method 2. Obviously, by replacing the derivative of the log-response probability by a simulator for the derivative of the response probability, some information is lost and \( \frac{\partial L^3}{\partial \theta} \) and the second component of \( \frac{\partial L^2}{\partial \theta} \) become indistinguishable. Method 2 has the intuitively appealing property that it includes the minimization of the distance between the binary response indicator \( d_n \) and its theoretical expectation \( P_n \). The simulation residual for method 3 can be found in the same way as for method 2 and is given by

\[ \text{RES}_3 = \frac{1}{R} \sum_{n=1}^{N} \left[ m_{nr} \frac{\partial P_n}{\partial \theta} - \frac{\partial \ln P_n}{\partial \theta} \right] - \sum_{n=1}^{N} v_n \frac{\partial \ln P_n}{\partial \theta} \]

The residual contains the same error which is not influenced by the number of drawings \( R \) as method 2.
3.4. Asymptotic Distribution of the Estimators

In the preceding subsections we presented three ways to simulate the first order derivatives of the log-likelihood function. The simulated score vectors of methods 2 and 3 satisfy the property that their expectation, evaluated at the true parameter vector \( \theta_0 \), equals zero, whereas the probability limit of the simulated score of method 1 equals zero at the true parameter value \( \theta_0 \) if the vectors of instruments are constructed in a proper way. Therefore, the length of the expectation of the simulated score is minimized at the true parameter value \( \theta_0 \). It is intuitively clear that if we minimize the length of the simulated score, the resulting parameter vector \( \hat{\theta}_R \) at which minimization takes place, will converge to the true parameter value \( \theta_0 \), or, equivalently, \( \hat{\theta}_R \) will be a consistent estimator of \( \theta_0 \).

We assume that

\[
\frac{1}{\sqrt{N}} K_R^i(\theta_0) \sim \mathcal{N}(0, V_{R_0}^i), \quad i = 1, \ldots, 3
\]

with \( V_{R_0}^i \) some positive definite symmetric matrix. Below, we explain that this assumption can be justified on the basis of central limit arguments. Using this assumption and the consistency of \( \hat{\theta}_R \), apart from usual regularity assumptions on the parameter space and the like, it is possible to show (see, for example, Pakes and Pollard [1989]) that

\[
\sqrt{N}(\hat{\theta}_R - \theta_0) \sim \mathcal{N}(0, (\Gamma'\Gamma)^{-1} \Gamma' V_{R_0}^i \Gamma (\Gamma'\Gamma)^{-1})
\]

where

\[
\Gamma' = \lim_{N \to \infty} \frac{1}{N} \left( \frac{\partial L'(\theta_0)/\partial \theta'}{\partial \theta} \right)
\]

We now will comment on the assumed normality of the simulated score. The expectation of the simulated score, conditional on the instruments, equals zero at \( \theta_0 \) for \( i = 2, 3 \). It was shown that the simulated score could be written as the sum of the true score vector and a simulation residual. It is well-known that under general conditions the distribution of the true score vector divided by the square root of the sample size converges to a normal distribution. The assumed independence across observations and the fact that the simulators are constructed using independent random drawings can be used to apply the Lindeberg-Feller central limit theorem to prove the normality of the simulation residuals (conditional on the instruments) divided by the square root of the sample size. The distribution of the simulated score then converges to the sum of two normal distributions which is in turn a normal distribution. We can do the same for method 1, but not without recalling that the instruments have to be constructed such that (19) is satisfied.

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

\[ \sqrt{N} (\hat{\theta}_{ML} - \theta_0) \sim \text{N}(0, \Omega_{ML}) \]

where

\[ \Omega_{ML} = B^{-1} \]

\[ B = -p \lim \frac{1}{N} \frac{\partial^2 L(\theta_0)}{\partial \theta \partial \theta'} \]

To make clear the relation with the asymptotic covariance matrix of the simulation estimators we rewrite \( \Omega_{ML} \) as

\[ \Omega_{ML} = (\Gamma'_{ML} \Gamma_{ML})^{-1} \Gamma'_{ML} V_{ML} \Gamma_{ML} (\Gamma'_{ML} \Gamma_{ML})^{-1} \]

where

\[ \Gamma'_{ML} = p \lim \frac{1}{N} \frac{\partial (\partial L(\theta_0)/\partial \theta')}{\partial \theta} = -B \]

which is the equivalent of (53), and

\[ V_{ML} = p \lim \frac{1}{N} \sum_{n=1}^{N} \frac{\partial L_n(\theta_0)}{\partial \theta} \frac{\partial L_n(\theta_0)}{\partial \theta'} \]

Recall that \( V_{ML} = B \). For method 1, if the instruments are constructed properly and if the number of drawings to construct the instruments tend to infinity \( \Gamma' \) and \( \Gamma'_{ML} \) are equivalent. Then the efficiency comparison reduces to comparing \( V_{ML} \) with \( V_{R} \) for this method. The difference between these matrices is given by the covariance matrix of the simulation residuals. In section 3.1 it has been derived that this variance disappears if \( R \) tends to infinity. Therefore we can conclude that if the matrix of instruments is constructed on the basis of (28) and if both the number of drawings to construct this matrix and the number of drawings to simulate the response probabilities tend to infinity, the covariance matrix of the method 1 estimator and the covariance matrix of the maximum likelihood estimator are asymptotically equal. Of course, this result has only theoretical meaning because the reason why we construct these simulation estimators is to be able to keep the number of drawings fixed and small.

To examine the efficiency of method 2 it has to be noted first that because of replacement (36), the simulated score \( K^2_R(\theta) \) does not tend to the true score if \( N \) is fixed and \( R \) tends to infinity. Therefore, we first need to establish the relation between \( \Gamma'_{ML} \) and \( \Gamma^{2'} \). From (33) and (37) it is readily
established that

\[
\Gamma^{2'} - \Gamma'_{ML} = -p \lim_{N} \sum_{n=1}^{N} \left[ (Z_n - Z_n) \frac{\partial P_n}{\partial \theta} + \frac{\partial Z'}{\partial \theta} (d_n - P_n) + v_n \frac{\partial (\partial \ln P_n / \partial \theta')}{} + \frac{1}{P} \frac{\partial P_n}{\partial \theta} \frac{\partial P_n}{\partial \theta'} \right]
\]

from which only the first three terms equal zero if the instruments are constructed such that (19) is satisfied, i.e. according to formula (28) with drawings tending to infinity. From the analysis of the simulation residuals it becomes clear that if the matrix of instruments is constructed according to (28) with drawings tending to infinity, and if the response probabilities and their derivatives are simulated with \( R \) tending to infinity as well, the asymptotic variance of the score of the likelihood function, evaluated in a consistent estimator is exceeded by \( X \), where

\[
X = p \lim \left( \frac{1}{N} \sum_{n=1}^{N} P_n P_n \frac{\partial \ln P_n}{\partial \theta} \frac{\partial \ln P_n}{\partial \theta'} \right)
\]

which was derived in section 3.2. The same expression can be derived for method 3.

Finally, to estimate the covariance matrix we calculate

\[
\hat{\Omega}_R = (\hat{f}^{ii})^{-1} \hat{f}^{ii} \hat{V}^i_R \hat{f}^i (\hat{f}^{ii})^{-1}
\]

with

\[
\hat{f}^{ii} = \frac{1}{N} \frac{\partial (\partial L'(\hat{\theta}_k))/\partial \theta'}{}
\]

\[
\hat{V}^i_R = \frac{1}{N} \sum_{n=1}^{N} K^i_{nR} (\hat{\theta}_{R}) K^i_{nR} (\hat{\theta}^i_{R})'
\]

where the index \( n \) indicates the \( n \)-th component of the simulated score. Expression (64) can be calculated by simulation.

4 Monte Carlo and Empirical Application

We will now illustrate the properties of the various estimators by making specific assumptions about the form of the labour supply function and the log-wage equation. We then compare the estimation methods first by Monte Carlo methods and next by estimating the specification for a sample
of 849 married female individuals, drawn in 1985. We'll assume that the preferences of the individuals are described by a utility function introduced by Hausman and Ruud [1984], which implies a labour supply function quadratic in the wage rate and linear in non-labour income. The wage equation is assumed to be log-linear. The disturbances of the labour supply function and the log-wage equation are assumed to be normally distributed.

The specific form of the labour supply function under the assumption of a linear budget constraint becomes:

\[ h(w_n, \mu_n, \beta) = \beta_3 + \mu_n^* \beta_2 + w_n \beta_4 \]

with

\[ \mu_n^* = \beta_1 + \mu_n + w_n \beta_3 + \frac{1}{2} w_n^2 \beta_4 \]

Inserting the expression for \( \mu_n^* \) in the labour supply function and reparameterizing gives:

\[ h(w_n, \mu_n, \alpha) = \alpha_1 + \mu_n \alpha_2 + w_n \alpha_3 + \frac{1}{2} w_n^2 \alpha_4 \]

with

\[ \alpha_1 = \beta_3 + \beta_1 \beta_2 \]
\[ \alpha_2 = \beta_2 \]
\[ \alpha_3 = \beta_2 \beta_3 + \beta_4 \]
\[ \alpha_4 = \beta_2 \beta_4 \]

The log-wage equation becomes:

\[ \log w_n = \sum_{j=1}^{11} \eta_j x_{nj} + u_n \]

where

- \( x_{n1} = 1 \) for all \( n \),
- \( x_{n2} = \log \) of number of persons in individual \( n \)'s family,
- \( x_{n3} = \) the number of children with age below 6 of individual \( n \),
- \( x_{n4} = \log \)-age of individual \( n \),
- \( x_{n5} \) to \( x_{n8} \) are dummy indicators for the level of education of individual \( n \), where \( x_{n5} \) is the lowest level of education, \( x_{n6} \) is the next to the lowest education level etc. For the highest education level no dummy indicator is included
- \( x_{n9} \) and \( x_{n10} \) are indicators for the type of education received, \( x_{n9} \) is a dummy indicator for non-technical and non-commercial type of education, \( x_{n10} \) is a dummy indicator for semi-technical and semi-commercial type of education.
For technical and commercial type of education no dummy indicator is included,

- \( x_{n11} = x_{n4}^2 \)
- here \( w_n \) is the after tax hourly wage rate; labour supply will be measured in hours per week

The complete model thus reads

\[
\begin{align*}
\text{(69)} & \quad h_n^* = \alpha_1 + \mu_n \alpha_2 + w_n \alpha_3 + \frac{1}{2} w_n^2 \alpha_4 + \epsilon_n \\
\text{(70)} & \quad \log w_n = \sum_{j=1}^{11} \eta_j x_{nj} + u_n
\end{align*}
\]

\( \epsilon_n \) and \( u_n \) are jointly normally distributed with

\[
\begin{align*}
\text{(71)} & \quad \begin{pmatrix} \epsilon_n \\ u_n \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma \right) \\
\text{(72)} & \quad \Sigma = \begin{pmatrix} \sigma_e^2 & \sigma_{eu} \\ \sigma_{eu} & \sigma_u^2 \end{pmatrix} \\
\text{(73)} & \quad h_n = h_n^* \quad \text{if } h_n^* > 0 \\
\text{(74)} & \quad h_n = 0 \quad \text{if } h_n^* \leq 0
\end{align*}
\]

We now derive an expression for the response probabilities and their derivatives. Under the normality assumptions the joint probability density function of \( h_n \) and \( w_n \) is

\[
\begin{align*}
\text{(75)} & \quad g \left( h_n, w_n \mid x_n, \mu_n, \theta \right) = P_n(\theta) \quad \text{if } h_n = 0 \\
& \quad \frac{1}{2 \pi w_n} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} \begin{pmatrix} h_n - h(w_n, \mu_n, \alpha) \end{pmatrix} \cdot \Sigma^{-1} \begin{pmatrix} h_n - h(w_n, \mu_n, \alpha) \end{pmatrix} \right\} \\
& \quad \text{if } h_n > 0, 0 < w_n < \infty
\end{align*}
\]

where

\[
\begin{align*}
\text{(76)} & \quad P_n(\theta) = \int_{-\infty}^{\infty} \Phi \left( \begin{pmatrix} \frac{h \left[ \exp (\eta' x_n + \sigma_u v) \right], \mu_n, \alpha} \sigma_{e|u} \end{pmatrix} \right. \\
& \quad \times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} v^2 \right\} \, dv
\end{align*}
\]

with

\[
\rho = \frac{\sigma_{eu}}{\sigma_e \sigma_u}, \quad \sigma_{e|u}^2 = \sigma_e^2 (1 - \rho^2)
\]

and where \( \Phi(\cdot) \) is the standard normal distribution function.

We have written the double integral in a form with the well-known function \( \Phi(\cdot) \) and an integral over the standard normal variable \( v \). As a result we only need to simulate the second integral over \( v \). Because the
standard normal distribution function doesn't depend on the parameters, we don't have to choose a weighting density function as described for the general case in the preceding section. To simulate the probability $P_n$ we can simply draw from the standard normal distribution and compute the expression under the integral sign.

We obtain the following expression for the smooth unbiased simulator of $P_n$.

$$k_n(0, v^*_n) = \frac{1}{R} \sum_{r=1}^{R} \Phi \left( -\frac{h[\exp(\eta' x_n + \sigma_u v^*_n), \mu_n, \alpha] + \rho \sigma_{x u} v^*_n}{\sigma_{e u}} \right)$$

where the $v^*_n$ are independent random drawings from the standard normal distribution.

To simulate the derivate of $P_n$ with respect to, say, $\theta_j$ we first write

$$\frac{\partial P_n}{\partial \theta_j} = \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta_j} \Phi \left( -\frac{h[\exp(\eta' x_n + \sigma_u v), \mu_n, \alpha] + \rho \sigma_{x u} v}{\sigma_{e u}} \right) dv$$

where the derivative in the integrand can be calculated analytically. Then we simulate it by

$$m_{nj}(0, v^*_n) = \frac{1}{R} \sum_{r=1}^{R} \frac{\partial}{\partial \theta_j} \Phi \left( -\frac{h[\exp(\eta' x_n + \sigma_u v^*_n), \mu_n, \alpha] + \rho \sigma_{x u} v^*_n}{\sigma_{e u}} \right)$$

4.1. Monte Carlo Results

To get an idea about the performance of the estimation methods we have performed some Monte Carlo experiments. First of all, values for the parameters in (69), (70), (71) and (78) were chosen. Next, disturbances $e_n$ and $u_n$, $n = 1, \ldots, N$, were generated, under the assumptions (71), (72) with $\sigma_{e u} = 0$. The characteristics $x_{nj}$ from the sample were used to generate wages $w_n$, $n = 1, \ldots, N$. Included are the constant term with parameter $\eta_1$, log(age) with parameter $\eta_2$ and the square of log(age) with parameter $\eta_3$. These wages were used to generate the $h^*_n$, $n = 1, \ldots, N$. Making use of (73) and (74) and the non-labour income series from the sample, we generated the “observations” $h_n$.

This generated data set was used to estimate the model with different methods of estimation. First of all the model is estimated with maximum likelihood (ML) using numerical integration. The Gauss-Hermite quadrature formula is used with a number of abscissae equal to 16, see e.g. STRoud and SecRest [1966]. The second method of estimation is simulated maximum likelihood (SML) with a number of drawings $R$ equal to 10. Here the probabilities are simulated according to (78) and inserted in the likelihood function directly. Although this method is inconsistent for...
$N$ is the number of parameters, whereas the subscript $j$ stands for the $j$-th component of the parameter vector. The second line gives the sample standard deviation of the estimates:

\[
SD_j = \sqrt{\frac{1}{20-1} \sum_{i=1}^{20} (\hat{\theta}_{ij} - \bar{\theta}_j)^2}, \quad j = 1, \ldots, N
\]

The third line presents the relative error:

\[
\text{rel. err.} = 100\% \times \frac{\| \hat{\theta}_j - \theta_{0j} \|}{\| \theta_{0j} \|}
\]

From the table it can be seen that the ML estimator in general performs best in the sense that it has the lowest standard errors and the lowest relative errors. It is clear that SML with $R = 10$ performs very badly. It performs even worse than all of the three MSS variants with $R = 1$. Therefore, the use of MSS instead of SML is not just something which only has theoretical relevance. Comparing different numbers of drawings of the same MSS estimator it can be seen that the standard deviations decrease with the increase in number of drawings, which is consistent with the analysis of the simulation residuals in chapter 3. For MSS1 the number of drawings $R_Z$ to simulate the matrix of instruments has been increased from 10 to 500. Recall that for this method the construction of the matrix of instruments not only affects the efficiency of the estimator but also its consistency. MSS1 with $R = 1$ and $R_Z = 500$ outperforms MSS1 with $R = 10$ and $R_Z = 10$. The standard deviations are comparable with those of MSS2 with $R = 1$ and $R_Z = 10$, but they are not lower than the MSS2 standard deviations, which questions the use of MSS1. Comparing MSS2 with MSS3 it can be said that for $R = 1$ method 2 has the lower standard deviations and the higher relative errors, whereas for $R = 10$ the differences are rather small, although MSS3 seems to perform slightly better, which may be due to the fact in MSS2 theoretically an additional inefficiency is introduced by simulating the matrix of instruments with only $R_Z = 10$ drawings.

One may question the practical relevance of the experiments so far with respect to MSS1 and MSS2, since the instruments were computed at the true parameter point which of course is unknown in practice. To see how this affects results, Table 4.2 presents results for MSS1 and MSS2 with instruments based on estimates obtained with MSS3.

Moreover, Table 4.2 also presents the mean standard error of the estimates over the twenty replications so that a comparison with the sample standard deviations is possible. Comparing the table with the corresponding columns in Table 4.1, one observes that MSS1 is a bit more sensitive to the choice of instruments than MSS2, as one would expect. Actually, MSS2 is hardly affected at all by the new instruments. Although MSS1 does give somewhat different estimates now, they are not systematically worse than before, sometimes the relative error is better, and sometimes it is worse. It remains true that the small value for $R_Z$ induces inefficiency. Finally, we observe that the estimated standard errors tend to be of a similar magnitude as the standard deviations, though with some
<table>
<thead>
<tr>
<th>Parameter</th>
<th>MSS1 ( R = 10, ) ( R_z = 10 )</th>
<th>MSS1 ( R = 1, ) ( R_z = 500 )</th>
<th>MSS2 ( R = 1, ) ( R_z = 10 )</th>
<th>MSS2 ( R = 10, ) ( R_z = 10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_1 )</td>
<td>-22.589</td>
<td>-10.765</td>
<td>-2.517</td>
<td>-6.705</td>
</tr>
<tr>
<td>SD</td>
<td>11.920</td>
<td>6.556</td>
<td>6.880</td>
<td>3.407</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>160.1</td>
<td>23.9</td>
<td>71.0</td>
<td>22.8</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>-0.0436</td>
<td>-0.0490</td>
<td>-0.0493</td>
<td>-0.0503</td>
</tr>
<tr>
<td>SD</td>
<td>0.0185</td>
<td>0.0136</td>
<td>0.0110</td>
<td>0.0116</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>9.451</td>
<td>1.752</td>
<td>2.405</td>
<td>4.388</td>
</tr>
<tr>
<td>( \alpha_3 )</td>
<td>4.535</td>
<td>3.353</td>
<td>2.624</td>
<td>2.921</td>
</tr>
<tr>
<td>SD</td>
<td>1.220</td>
<td>0.685</td>
<td>0.528</td>
<td>0.303</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>44.6</td>
<td>6.890</td>
<td>16.4</td>
<td>6.890</td>
</tr>
<tr>
<td>( \alpha_4 )</td>
<td>0.221</td>
<td>-0.173</td>
<td>-0.147</td>
<td>-0.156</td>
</tr>
<tr>
<td>SD</td>
<td>3.0548</td>
<td>0.0300</td>
<td>0.0194</td>
<td>0.0139</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>35.6</td>
<td>5.976</td>
<td>9.913</td>
<td>4.340</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>366.357</td>
<td>349.480</td>
<td>347.272</td>
<td>347.013</td>
</tr>
<tr>
<td>SD</td>
<td>63.981</td>
<td>8.146</td>
<td>0.469</td>
<td>0.163</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>5.602</td>
<td>0.736</td>
<td>0.101</td>
<td>0.0266</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>7.586</td>
<td>2.809</td>
<td>0.995</td>
<td>1.004</td>
</tr>
<tr>
<td>SD</td>
<td>16.004</td>
<td>6.503</td>
<td>0.0869</td>
<td>0.0686</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>602.4</td>
<td>160.1</td>
<td>7.896</td>
<td>7.066</td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>-8.430</td>
<td>-8.694</td>
<td>-12.602</td>
<td>-12.022</td>
</tr>
<tr>
<td>SD</td>
<td>3.342</td>
<td>5.512</td>
<td>5.382</td>
<td>0.630</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>28.6</td>
<td>26.3</td>
<td>6.783</td>
<td>1.871</td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>-7.299</td>
<td>-7.100</td>
<td>9.215</td>
<td>8.859</td>
</tr>
<tr>
<td>SD</td>
<td>1.683</td>
<td>2.963</td>
<td>2.970</td>
<td>0.389</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>15.6</td>
<td>17.9</td>
<td>6.609</td>
<td>2.482</td>
</tr>
<tr>
<td>( \eta_3 )</td>
<td>-1.003</td>
<td>-0.990</td>
<td>-1.291</td>
<td>-1.243</td>
</tr>
<tr>
<td>SD</td>
<td>0.227</td>
<td>0.423</td>
<td>0.411</td>
<td>0.0689</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>16.4</td>
<td>17.4</td>
<td>7.707</td>
<td>3.658</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MSS3 ( R = 1 )</th>
<th>MSS3 ( R = 10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_1 )</td>
<td>-11.786</td>
<td>-8.690</td>
</tr>
<tr>
<td>SD</td>
<td>8.002</td>
<td>0.0136</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>35.7</td>
<td>0.00420</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>-0.0445</td>
<td>-0.0506</td>
</tr>
<tr>
<td>SD</td>
<td>0.0193</td>
<td>0.141</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>7.609</td>
<td>5.125</td>
</tr>
<tr>
<td>( \alpha_3 )</td>
<td>3.305</td>
<td>3.091</td>
</tr>
<tr>
<td>SD</td>
<td>0.641</td>
<td>0.178</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>3.355</td>
<td>1.460</td>
</tr>
<tr>
<td>( \alpha_4 )</td>
<td>-0.168</td>
<td>-0.161</td>
</tr>
<tr>
<td>SD</td>
<td>0.0230</td>
<td>0.0100</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>3.233</td>
<td>0.982</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>342.439</td>
<td>346.921</td>
</tr>
<tr>
<td>SD</td>
<td>19.954</td>
<td>0.000762</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>1.292</td>
<td>0.0475×10⁻³</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>1.116</td>
<td>1.065</td>
</tr>
<tr>
<td>SD</td>
<td>0.155</td>
<td>0.0994</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>3.332</td>
<td>1.427</td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>-12.031</td>
<td>-11.775</td>
</tr>
<tr>
<td>SD</td>
<td>7.450</td>
<td>0.9555</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>1.946</td>
<td>0.225</td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>8.651</td>
<td>8.696</td>
</tr>
<tr>
<td>SD</td>
<td>4.092</td>
<td>0.167</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>0.0850</td>
<td>0.599</td>
</tr>
<tr>
<td>( \eta_3 )</td>
<td>-1.185</td>
<td>-1.216</td>
</tr>
<tr>
<td>SD</td>
<td>0.565</td>
<td>0.0544</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>1.127</td>
<td>1.418</td>
</tr>
</tbody>
</table>

NON-LINEAR LABOUR SUPPLY FUNCTION AND WAGE EQUATION 195
a fixed and small number of drawings, it still would be useful if the
asymptotic bias were small. Finally, the model is estimated using the three
MSS estimators, abbreviated as MSS1, MSS2 and MSS3 below. Two
different numbers of drawings to simulate the response probabilities are
used, i.e. R = 1 and R = 10. The matrix of instruments is constructed on
the basis of formula (28) evaluated in the true parameter-point, where the
number of drawings to simulate the instruments is R = 10 for both MSS1
and MSS2. For MSS1 the model is also estimated with R = 500 and
R = 1.

Tables 4.1 presents the results of the Monte Carlo study. The Monte
Carlo procedure was repeated 20 times, so that the numbers in the tables
refer to means over 20 replications. This modest number of replications
has been chosen in view of the rather heavy computational burden of non-
linear optimization problems in general. For each parameter the first line
presents the mean of the estimates as given by:

\[ \bar{\theta}_j = \frac{1}{20} \sum_{i=1}^{20} \hat{\theta}_{ij}, \quad j = 1, \ldots, N \]

**Table 4.1**

**Monte Carlo Results**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>ML ( R = 10 )</th>
<th>SML ( R = 10 )</th>
<th>MSS1 ( R = 1 ), MSS2 ( R = 10 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_1 )</td>
<td>-8.686</td>
<td>-9.961</td>
<td>-5.996</td>
</tr>
<tr>
<td>SD</td>
<td>0.586</td>
<td>15.689</td>
<td>12.004</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>3.166</td>
<td>31.0</td>
<td>120.1</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-0.0482</td>
<td>-0.0483</td>
<td>-0.0414</td>
</tr>
<tr>
<td>SD</td>
<td>0.0146</td>
<td>0.653</td>
<td>0.0126</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>0.265</td>
<td>760.4</td>
<td>4.443</td>
</tr>
<tr>
<td>( \alpha_3 )</td>
<td>3.137</td>
<td>3.098</td>
<td>-2.258</td>
</tr>
<tr>
<td>SD</td>
<td>0.0784</td>
<td>10.249</td>
<td>1.118</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>1.259</td>
<td>172.0</td>
<td>30.4</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>-0.163</td>
<td>-0.156</td>
<td>-0.109</td>
</tr>
<tr>
<td>SD</td>
<td>0.00850</td>
<td>0.0598</td>
<td>0.0587</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>4.518</td>
<td>33.2</td>
<td>29.0</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>346.921</td>
<td>346.923</td>
<td>254.785</td>
</tr>
<tr>
<td>SD</td>
<td>0.0233</td>
<td>255.331</td>
<td>581.318</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>0.00552</td>
<td>26.6</td>
<td>73.4</td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>1.080</td>
<td>1.175</td>
<td>0.448</td>
</tr>
<tr>
<td>SD</td>
<td>0.123</td>
<td>0.572</td>
<td>2.311</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>8.809</td>
<td>58.6</td>
<td>82.0</td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>-11.801</td>
<td>-11.860</td>
<td>158.243</td>
</tr>
<tr>
<td>SD</td>
<td>0.0507</td>
<td>202.316</td>
<td>5.900</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>0.501</td>
<td>1440.1</td>
<td>23.9</td>
</tr>
<tr>
<td>( \eta_3 )</td>
<td>8.644</td>
<td>8.560</td>
<td>5.600</td>
</tr>
<tr>
<td>SD</td>
<td>0.118</td>
<td>6.436</td>
<td>3.284</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>0.972</td>
<td>35.2</td>
<td>17.6</td>
</tr>
<tr>
<td>( \eta_4 )</td>
<td>-1.199</td>
<td>-1.189</td>
<td>-3.801</td>
</tr>
<tr>
<td>SD</td>
<td>0.0383</td>
<td>7.795</td>
<td>0.453</td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>0.844</td>
<td>217.0</td>
<td>18.4</td>
</tr>
</tbody>
</table>

196
exceptions. The standard errors tend to be a little higher than the standard deviations, hence one could take standard errors as a somewhat conservative estimate of the inaccuracy of the estimates.

We conclude that for SML the limited number of drawings of 10 is clearly not sufficient for a reasonable performance of the estimator. All of the three MSS estimators perform acceptably even with only one drawing, although it is clear that efficiency can be improved by taking more than one drawing.

**Table 4.2**

**Additional Monte Carlo Results, Two Stages**

<table>
<thead>
<tr>
<th>parameter</th>
<th>MSS1</th>
<th></th>
<th>MSS2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R = 10, )</td>
<td>( R = 10, )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( R_x = 10 )</td>
<td>( R_x = 10 )</td>
<td></td>
</tr>
<tr>
<td>( a_1 )</td>
<td>-19.261</td>
<td>-6.146</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>11.771</td>
<td>4.086</td>
<td></td>
</tr>
<tr>
<td>Est. SE</td>
<td>14.623</td>
<td>9.369</td>
<td></td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>121.8</td>
<td>29.242</td>
<td></td>
</tr>
<tr>
<td>( a_2 )</td>
<td>-0.0560</td>
<td>-0.0508</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.0177</td>
<td>0.0121</td>
<td></td>
</tr>
<tr>
<td>Est. SE</td>
<td>0.0106</td>
<td>0.00989</td>
<td></td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>0.068</td>
<td>0.2960</td>
<td></td>
</tr>
<tr>
<td>( a_3 )</td>
<td>4.068</td>
<td>2.902</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>1.073</td>
<td>0.366</td>
<td></td>
</tr>
<tr>
<td>Est. SE</td>
<td>1.318</td>
<td>0.607</td>
<td></td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>29.678</td>
<td>7.490</td>
<td></td>
</tr>
<tr>
<td>( a_4 )</td>
<td>-0.211</td>
<td>-0.156</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.0528</td>
<td>0.0154</td>
<td></td>
</tr>
<tr>
<td>Est. SE</td>
<td>0.0674</td>
<td>0.0231</td>
<td></td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>29.155</td>
<td>4.547</td>
<td></td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>647.011</td>
<td>347.050</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>460.833</td>
<td>0.210</td>
<td></td>
</tr>
<tr>
<td>Est. SE</td>
<td>493.617</td>
<td>103.994</td>
<td></td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>86.501</td>
<td>0.371 ( \times 10^{-3} )</td>
<td></td>
</tr>
<tr>
<td>( \eta_1 )</td>
<td>5.095</td>
<td>1.0001</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>12.633</td>
<td>0.0756</td>
<td></td>
</tr>
<tr>
<td>Est. SE</td>
<td>2.673</td>
<td>0.0696</td>
<td></td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>371.8</td>
<td>7.335</td>
<td></td>
</tr>
<tr>
<td>( \eta_2 )</td>
<td>-11.286</td>
<td>-11.650</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>12.470</td>
<td>10.542</td>
<td></td>
</tr>
<tr>
<td>Est. SE</td>
<td>14.975</td>
<td>11.945</td>
<td></td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>4.371</td>
<td>0.152</td>
<td></td>
</tr>
<tr>
<td>( \eta_3 )</td>
<td>8.814</td>
<td>8.647</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>7.301</td>
<td>4.928</td>
<td></td>
</tr>
<tr>
<td>Est. SE</td>
<td>8.539</td>
<td>6.734</td>
<td></td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>1.970</td>
<td>0.299 ( \times 10^{-3} )</td>
<td></td>
</tr>
<tr>
<td>( \eta_4 )</td>
<td>-1.244</td>
<td>-1.212</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>1.020</td>
<td>0.843</td>
<td></td>
</tr>
<tr>
<td>Est. SE</td>
<td>1.191</td>
<td>0.951</td>
<td></td>
</tr>
<tr>
<td>rel. err. (%)</td>
<td>2.119</td>
<td>1.098</td>
<td></td>
</tr>
</tbody>
</table>
4.2. Estimation Results

We now present the estimates of the model for the real data. The model has been estimated using the three methods described in section 3 and by ML. For each method, the model was estimated under the assumption \( \sigma_{u} = 0 \) with two different numbers of drawings; \( R = 5 \) and \( R = 10 \). The restriction \( \sigma_{u} = 0 \) is relaxed below. The instrument matrices are constructed using \( R_{Z} = 10 \) drawings.

In table 4.3 the ML estimates are given. The estimates have the expected sign. The maximum number of hours worked occurs at a wage rate of about 11 guilders per hour, the wage rate reaches its maximum at the age of 35 years.

**Table 4.3**

*Maximum Likelihood Estimates*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \hat{\theta} )</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_{1} )</td>
<td>-86.696</td>
<td>10.860</td>
</tr>
<tr>
<td>( \alpha_{2} )</td>
<td>-5.605 \times 10^{-2}</td>
<td>0.00936</td>
</tr>
<tr>
<td>( \alpha_{3} )</td>
<td>12.112</td>
<td>1.312</td>
</tr>
<tr>
<td>( \alpha_{4} )</td>
<td>-0.574</td>
<td>0.0625</td>
</tr>
<tr>
<td>( \sigma_{1} )</td>
<td>558.052</td>
<td>96.753</td>
</tr>
<tr>
<td>( \sigma_{2} )</td>
<td>0.135</td>
<td>0.0164</td>
</tr>
<tr>
<td>( \eta_{1} )</td>
<td>-14.959</td>
<td>3.733</td>
</tr>
<tr>
<td>( \eta_{2} )</td>
<td>-0.387</td>
<td>0.781</td>
</tr>
<tr>
<td>( \eta_{3} )</td>
<td>-8.765 \times 10^{-2}</td>
<td>0.0206</td>
</tr>
<tr>
<td>( \eta_{4} )</td>
<td>10.322</td>
<td>2.110</td>
</tr>
<tr>
<td>( \eta_{5} )</td>
<td>-0.767</td>
<td>0.119</td>
</tr>
<tr>
<td>( \eta_{6} )</td>
<td>-0.771</td>
<td>0.117</td>
</tr>
<tr>
<td>( \eta_{7} )</td>
<td>-0.625</td>
<td>0.115</td>
</tr>
<tr>
<td>( \eta_{8} )</td>
<td>-0.317</td>
<td>0.112</td>
</tr>
<tr>
<td>( \eta_{9} )</td>
<td>0.266</td>
<td>0.410</td>
</tr>
<tr>
<td>( \eta_{10} )</td>
<td>5.611 \times 10^{-2}</td>
<td>0.0511</td>
</tr>
<tr>
<td>( \eta_{11} )</td>
<td>-1.452</td>
<td>0.295</td>
</tr>
</tbody>
</table>

In table 4.4 the estimates by method 1, using \( R = 5 \) drawings from the standard normal distribution, and their asymptotic standard errors are presented. To circumvent problems with consistency, we used the estimates of method 3 (see Table 4.9) to construct the matrix of instruments. Most of the parameter estimates have the expected sign. Non-labour income has a negative impact on labour supply. The estimate of the linear wage parameter is positive and the estimate of the squared wage parameter is negative. We can calculate that the number of hours worked reaches its maximum at a wage rate of Dfl. 13.5 per hour. Log-family size influences the log-wage rate negatively. The education dummies have the expected negative sign. Moreover, the higher the level of education, the less negative is the parameter estimate of the corresponding education dummy. From the parameter estimates of log-age and its square we can calculate that the wage rate reaches its maximum at the age of 38.
### Table 4.4

**Estimates by Method 1 Number of Drawings R = 5**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\hat{\delta}^1$</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1$</td>
<td>-22.595</td>
<td>9.43</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$-6.194 \times 10^{-2}$</td>
<td>0.02</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>3.158</td>
<td>1.12</td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>0.117</td>
<td>0.05</td>
</tr>
<tr>
<td>$\sigma_1^2$</td>
<td>309.942</td>
<td>17.53</td>
</tr>
<tr>
<td>$\sigma_2^2$</td>
<td>$6.819 \times 10^{-2}$</td>
<td>0.01</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>-11.855</td>
<td>2.43</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>0.118</td>
<td>0.06</td>
</tr>
<tr>
<td>$\eta_3$</td>
<td>$1.353 \times 10^{-2}$</td>
<td>0.03</td>
</tr>
<tr>
<td>$\eta_4$</td>
<td>8.198</td>
<td>1.39</td>
</tr>
<tr>
<td>$\eta_5$</td>
<td>-0.522</td>
<td>0.10</td>
</tr>
<tr>
<td>$\eta_6$</td>
<td>-0.454</td>
<td>0.10</td>
</tr>
<tr>
<td>$\eta_7$</td>
<td>-0.395</td>
<td>0.10</td>
</tr>
<tr>
<td>$\eta_8$</td>
<td>-0.156</td>
<td>0.10</td>
</tr>
<tr>
<td>$\eta_9$</td>
<td>0.163</td>
<td>0.04</td>
</tr>
<tr>
<td>$\eta_{10}$</td>
<td>$6.796 \times 10^{-2}$</td>
<td>0.04</td>
</tr>
<tr>
<td>$\eta_{11}$</td>
<td>-1.128</td>
<td>0.02</td>
</tr>
</tbody>
</table>

The estimates in table 4.5 are also obtained by applying method 1, but now we have used $R = 10$ drawings to construct the simulators. Again, most of the estimates have the expected sign. The maximum of the number of hours worked with respect to the wage rate is reached at $w = 13.4$. The log-wage rate reaches its maximum with respect to age at 38 years. The estimates don't differ much from the ones in Table 4.4, but the standard errors are somewhat lower.

In table 4.6 we present the estimates obtained by method 2. The matrix of instruments is constructed using the estimates in Table 4.9. Apart from the other parameter estimates, now also the parameter estimate of $\eta_3$ (number of children below the age of 6) has the expected sign. The wage is maximal at the age of 37 and the number of hours supplied is maximal at a wage rate of 54.7.

Table 4.7 shows the method 2 estimates when $R = 10$ drawings are used. Again, the matrix of instruments is constructed using the estimates in Table 4.9. The wage equation is maximal at the age of 37, whereas labour supply is maximal at a wage rate of 32.9. The main difference between the method 1 estimates and the method 2 estimates are the parameter estimates of the labour supply function.

Finally, we look at the estimates obtained by method 3. In Table 4.8 the results with $R = 5$ drawings are presented. We can make the same remarks about the signs of the estimates as in the previous cases. Wage is maximal at the age of 38 and hours supplied are maximal at the wage rate of 9.4. To obtain the results in Table 4.9, we used $R = 10$ drawings to construct the simulators. Most of the standard errors of the estimates are...
lower than in Table 4.8, and whenever they aren’t lower, they are only slightly higher.

200
In Table 4.10 we present the estimation results with \( R = 50 \) drawings. Comparing Tables 4.4 and 4.8, we see that there is not much difference between the results with \( R = 10 \) and \( R = 50 \) drawings. Apparently \( R = 10 \) drawings are in this case sufficient to minimize the effect of the simulation residuals.

Comparing the three methods, we can say that method 1 is the cheapest in CPU-time because it doesn’t make use of simulators of the derivatives. Also, it appears to produce the smallest standard errors. However, it only makes sense to use this method when a consistent estimate is available to construct the matrix of instruments. Method 2 is the most expensive in CPU-time.

Table 4.11 gives the results of estimating the model without the restriction of zero correlation between the disturbances by method 3, using 10 drawings to simulate the response probabilities. The estimates of the disturbances’ variances and the covariance imply a correlation coefficient of \( \rho = 0.082 \) which is not significantly different from zero. A comparison of Tables 4.11 and 4.9 reveals no big shifts in the parameter estimate.

To get some more feeling for the differences in estimates across methods, we present elasticities of hours worked and of participation with respect to wages. These have been calculated as “aggregate” elasticities in the sense that all wages in the sample have been raised by 5% and then hours and participation probabilities have been predicted for every individual in the sample. The observed changes in the sample averages of these quantities are used to compute the elasticities. The results are given in Table 4.12.
TABLE 4.11

Estimates by Method 3 Number of Drawings R = 10 Correlated Disturbances

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\hat{\theta}^3$</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>-8.994</td>
<td>10.872</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$-6.592 \times 10^{-2}$</td>
<td>0.01</td>
</tr>
<tr>
<td>$a_3$</td>
<td>3.298</td>
<td>1.84</td>
</tr>
<tr>
<td>$a_4$</td>
<td>$-0.172$</td>
<td>0.10</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>308.869</td>
<td>91.15</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>$7.054 \times 10^{-2}$</td>
<td>0.01</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>$-11.728$</td>
<td>2.75</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>$-0.115$</td>
<td>0.07</td>
</tr>
<tr>
<td>$\eta_3$</td>
<td>$3.992 \times 10^{-2}$</td>
<td>0.04</td>
</tr>
<tr>
<td>$\eta_4$</td>
<td>8.093</td>
<td>1.57</td>
</tr>
<tr>
<td>$\eta_5$</td>
<td>$-0.522$</td>
<td>0.11</td>
</tr>
<tr>
<td>$\eta_6$</td>
<td>$-0.442$</td>
<td>0.11</td>
</tr>
<tr>
<td>$\eta_7$</td>
<td>$-0.390$</td>
<td>0.10</td>
</tr>
<tr>
<td>$\eta_8$</td>
<td>$-0.149$</td>
<td>0.11</td>
</tr>
<tr>
<td>$\eta_9$</td>
<td>0.169</td>
<td>0.05</td>
</tr>
<tr>
<td>$\eta_{10}$</td>
<td>$7.060 \times 10^{-2}$</td>
<td>0.04</td>
</tr>
<tr>
<td>$\eta_{11}$</td>
<td>$-1.111$</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Table 4.12

Wage Elasticities According to Different Estimation Methods, R = 10

<table>
<thead>
<tr>
<th></th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours</td>
<td>1.119</td>
<td>0.637</td>
<td>0.518</td>
<td>1.943</td>
</tr>
<tr>
<td>Participation</td>
<td>0.606</td>
<td>0.133</td>
<td>0.236</td>
<td>1.369</td>
</tr>
</tbody>
</table>

Strikingly, ML gives elasticities that are much larger than those implied by the other methods. Method 1 is most similar to ML in this respect. It is hard to interpret these differences. In principle they would call for specification tests. Given the simplistic nature of the model and the illustrative purposes of the estimation we abstain from a specification search. ¹

As a final comparison of estimation methods, we present in Table 4.13 the likelihood values corresponding to the estimates obtained by the various

---

¹ Presumably the most important problem with the present model is that it assumes that anyone who wants to work can do so (if $h^* > 0$, one works a positive amount of hours). This assumption is far too strong, see for instance Blundell, Ham and Meghir [1987] or Kapteyn and Woittiez [1989].
### Table 4.13

**Values of the Likelihood Function**

<table>
<thead>
<tr>
<th>Estimates from table:</th>
<th>Likelihood value</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3 (ML)</td>
<td>-2738.050</td>
</tr>
<tr>
<td>4.4 (method 1)</td>
<td>-2881.038</td>
</tr>
<tr>
<td>4.7 (method 2)</td>
<td>-2866.351</td>
</tr>
<tr>
<td>4.9 (method 3)</td>
<td>-3006.794</td>
</tr>
</tbody>
</table>

methods. We now see that MSS2 is closest to ML and MSS3 has the lowest likelihood value.

## 5 Conclusions

The main purpose of the paper has been to investigate the usefulness of MSS estimators in mixed discrete-continuous models, with a focus on the kind of model typically encountered in the analysis of labour supply. The experience in this paper appears to be that the estimators perform quite well. In the example considered, the MSS estimators do a little worse than ML, but in more complex situations ML would simply be infeasible. This is not only a matter of computing time, but also due to the fact that in certain situations it is impossible to write down analytically the probability of certain events, whereas the events can still be simulated. Estimation by simulation techniques then turns out to be a useful tool in the analysis of labour supply models, in the sense that these techniques enable us to estimate models which cannot, or only with great difficulty, be estimated with conventional methods like maximum likelihood or the method of moments.

The simulated scores methods presented in this paper perform satisfactory, even with a limited number of drawings. For the use of a limited number of drawings a price has to be paid in the form of a loss in efficiency, but this loss is modest. This is in stark contrast with the method of simulated maximum likelihood with a limited of drawings, which performs poorly.

The method of simulated scores will be the more useful, the higher is the dimension of integration in the evaluation of response probabilities in the likelihood function. In this context one may think of models of family labour supply with various sources of randomness. In this paper we have only used smooth simulators. In more complex models the use of frequency...
simulators cannot always be avoided. Their main drawback is their discontinuity in the parameters as a result of which conventional gradient based optimization procedures cannot be used. The downhill simplex method, employed by Bloemen and Kapteyn [1992] for instance, is quite time consuming. This disadvantage, however, is mitigated by the possibility to use a limited number of drawings in the method of simulated scores.

• References


Reprint Series, CentER, Tilburg University, The Netherlands:


No. 5  Th. ten Raa and F. van der Ploeg, A statistical approach to the problem of negatives in input-output analysis, *Economic Modelling*, vol. 6, no. 1, 1989, pp. 2 - 19.


No. 8  Th. van de Klundert and F. van der Ploeg, Wage rigidity and capital mobility in an optimizing model of a small open economy, *De Economist*, vol. 137, nr. 1, 1989, pp. 47 - 75.


<table>
<thead>
<tr>
<th>No.</th>
<th>Authors and Titles</th>
</tr>
</thead>
</table>


<table>
<thead>
<tr>
<th>No.</th>
<th>Author(s)</th>
<th>Title</th>
<th>Source</th>
</tr>
</thead>
</table>


<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Authors</th>
<th>Journal/Book</th>
<th>Page(s)</th>
</tr>
</thead>
</table>


