ELSEVIER

Journal of Econometrics 107 (2002) 345–374

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

A structural labour supply model with flexible preferences

Arthur van Soest[a,*], Marcel Das[b], Xiaodong Gong[c]

[a]*Department of Econometrics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, Netherlands*
[b]*CentERdata, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, Netherlands*
[c]*IZA, P.O. Box 7240, D-53072 Bonn, Germany, and Research School of Social Sciences, Australian National University, Canberra, ACT 0200, Australia*

Abstract

We show how non-parametric flexibility can be attained in a structural labour supply model that can be used to analyse all sorts of (non-linear) tax and benefits reforms. The direct utility function is approximated with a series expansion. For given length of the expansion, the model is estimated by smooth simulated maximum likelihood, using Dutch data on labour supply of married females. Estimates of own and cross wage elasticities and tax reform effects suggest that a series expansion of order two is enough. Monte Carlo simulations show that the estimator performs very well, unless there is measurement error in the hours variable. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: C14; H31; J220

Keywords: Labour supply; Tax reforms; Non-parametric estimation

1. Introduction

Non-parametric regression techniques are usually seen as a statistic device for data description and exploration. They are typically not used for estimating more complex models with the rich economic structure required for policy analysis. Models for policy analysis are often characterised by restrictive functional form assumptions, which are needed to make the econometrics tractable, but lack sound economic foundation.

An example is the extensive literature on neo-classical structural labour supply models. This goes back to Hausman's kinked budget constraint model, extended and applied in numerous studies. See, for example, Burtless and Hausman (1978), Hausman (1981,

* Corresponding author. Tel.: +31-13-4662028; fax: +31-13-4663280.
E-mail address: avas@kub.nl (A. van Soest).

PII: S0304-4076(01)00128-2

1985), and the surveys of Moffitt (1986, 1990) and Blundell and MaCurdy (1999). This model is one of the first limited dependent variable models in micro-econometrics where economic theory and econometric specification are directly linked. In most of its applications, a restrictive specification of preferences is used, particularly in case of non-convexities in the budget set. The main reason is that an analytic solution of the utility maximisation problem then requires explicit expressions for both the direct utility function and the labour supply function (or the indirect utility or expenditure function). Thus the rich economic structure of the model hampers flexibility of the specification.

In the case of piecewise linear convex budget constraints, this problem can be avoided. Using the specific search algorithm for the optimum for this case (see Blomquist, 1983), only the specification of the labour supply curve is needed, and more flexible specifications come within reach. Blomquist and Newey (1997) have exploited this idea to estimate the labour supply curve non-parametrically, approximating it by a series expansion. This is to our knowledge the only example in this field where economic theory and non-parametrics are combined into one econometric model. Still, this study remains close to the original Hausman framework, not allowing for, for example, non-convex budget sets, unobserved wages of non-workers, or joint decisions of two spouses. The main reason is that Blomquist and Newey use the labour supply function only, avoiding the need to specify the underlying direct utility function.

In this paper, we build the model around a flexible, non-parametric specification of the direct utility function. We thus introduce a structural labour supply model with a flexible specification of preferences that can be used for the analysis of all sorts of (non-linear) tax and benefits changes. Following Van Soest (1995), we replace the actual budget set by a finite number of points, and we approximate the utility maximisation problem by finding the best point in this finite set. We do not require that the tax and benefits system is piecewise linear or convex, and we show how fixed costs of working, unobserved wages, and non-participation can be incorporated. Moreover, we show how our model avoids the critique by MaCurdy et al. (1990) that coherency of the model implicitly limits the range of elasticities that can be obtained, implying that policy outcomes may be driven by constraints on the chosen model rather than by the data and the estimates. Thus our framework allows for a flexible specification of preferences but also for many other features of structural labour supply models that have been addressed in the literature.

In our framework, the direct utility function is approximated by a series expansion in hours and income. Observed individual or family characteristics are incorporated through one or more of the coefficients in the series approximation. Unobserved heterogeneity is allowed for by adding random terms to one of the parameters. GEV I errors are added to the utility values of all alternatives in the finite choice set. These can be seen as alternative specific errors in utility evaluation or as a smoothing device (as in Keane and Moffitt, 1998). The wage equation is estimated jointly with the labour supply model and the participation decision is analysed as part of the labour supply decision.

For given length of the series expansion, our model can be estimated by smooth simulated maximum likelihood. The log likelihood has its usual Kullback–Leibler information criterion interpretation, but, due to the similarity with a multinomial logit model, it can also be given an interpretation in terms of information theory, similar to the normalised entropy (Golan, 1988) and the information index (Soofi, 1992). We will choose the length of the series expansion by comparing information indexes, log likelihoods, and Akaike Information Criterion values.

The estimates can be used to compute labour supply elasticities or to analyse the effects of changes in taxes and benefits on participation and labour supply. The model is estimated using data on married females in the Netherlands, drawn from the 1995 wave of the Dutch Socio-Economic Panel. We will illustrate the usefulness of our results for policy analysis by studying the consequences of a recently proposed income tax reform. This reform involves changing the rules for transferring the tax-free allowance from wife to husband if the wife works few hours. It requires a model that can deal with the participation decision and with non-convex budget frontiers. Our model is particularly appropriate for this.

In statistical terms, the method we propose is similar to, for example, the technique of semi-non-parametric maximum likelihood introduced by Gallant and Nychka (1987). The idea of that approach is to get a flexible specification of either the distribution of the error terms or the systematic part of the equations (or both), using series approximations. Once the number of terms in the series approximations is fixed, ML is performed. The non-parametric feature is that the number of terms can become large; it increases to infinity with the number of observations, but at a slower rate. We focus on the practical application of these ideas, and not on the asymptotics of the estimator if the number of terms in the series approximation tends to infinity. Thus, formally, our hypothesis tests and standard errors are only valid under the assumption that the length of the series approximation is given, and that the utility function is perfectly captured using this given length. In this case, standard properties of (parametric) simulated maximum likelihood apply. We will then compare results for different lengths of the series expansion. This makes our approach similar to applications of the Gallant and Nychka (1987) framework, such as Gabler et al. (1993).

The difference between our framework and many other non-parametric models is that economic theory is used to impose structure on the model. A utility function is estimated, but labour supply, i.e. the outcome of utility maximisation, is observed. One wage rate (at most) is observed for each individual, but the complete budget set is needed to solve the utility maximisation problem. Economic theory does not determine the functional form of the utility function. It therefore seems natural to use a flexible, non-parametric specification of the utility function, retaining the economic structure of the model. We thus combine information in the data with two types of prior information: the non-parametric assumption of utility maximisation, and the limitation of the number of terms in the series, required due to the finite size of the sample.

The structure of the remainder of this paper is as follows. In Section 2, we introduce the model and discuss the advantages and drawbacks of the discrete approach compared to the continuous approach. Here we also discuss the interpretation of the likelihood in terms of information theory, which is more complicated than in the standard multinomial logit model, due to the unobserved heterogeneity in preferences and wages.

In Section 3, we apply the model to Dutch data on labour supply of married females. We focus on the sensitivity of the results for the chosen length of the series approximation. We look at labour supply elasticities and at the effects on participation and hours worked of a recently proposed tax reform. In Section 4 we discuss several Monte Carlo simulations to investigate the quality of our ML estimator, and its sensitivity to misspecification. Section 5 concludes.

2. Model

We present a static neo-classical structural labour supply model. The basic framework is similar to that of Van Soest (1995) and Gong and van Soest (2002). We describe the model for single decision-makers. The model for joint decision-makers such as married couples is a straightforward generalisation, but is not the focus of the current paper. In the application below, we will analyse labour supply of married females, conditional on hours worked by their husbands (and on their husbands' earnings), where the woman is the only decision-maker in the model.

2.1. Utility

The individual's utility depends on leisure (TE-h, where TE is the time endowment), and on total net income of the family (y). Net income is partly determined by the individual's own earnings, but can also contain spouse's earnings, asset income, child allowances, etc. We follow the majority of labour supply studies, in which 'leisure' is the aggregate of all other time uses except work. The time endowment is a common constant for all individuals, and will drop out of the polynomial expansions. As a consequence, it is equivalent to use a direct utility function with arguments y and h instead of y and TE-h.

The model would be consistent with utility maximisation in a life cycle framework with inter-temporally separable preferences if net income could be replaced by total expenditures (see Blundell and Walker, 1986). Our data do not contain any information on consumption expenditures or savings, so that we remain in a static framework.

The direct utility function is specified as a polynomial in its arguments h and y:

$$U(h,y) = \sum_{p=0,\dots,K} \sum_{q=0,\dots,K-p} \alpha(p,q) h^p y^q, \tag{1}$$

where K is the order of the polynomial and determines the flexibility of the utility function. If K can become arbitrarily large, the parameters $\alpha(p,q)$ can be chosen in such a way that $U(h,y)$ can approximate any given function of h and y to any desired accuracy on a given compact set. In this sense, the class of utility functions in (1) can be seen as a non-parametric family of utility functions. On the other hand, for a finite sample, the order of the polynomial that can be used is limited.

As usual with non-parametric series expansions, the proof of consistency of the estimator will require that K tends to infinity at a slower rate than the number of observations. In practice however, only small values of K can be used in estimation for the given size of the sample. We will use $K = 1, 2, 3, 4$ and 5. For $K = 5$, the approximation already contains 20 terms.[1]

The economic interpretation of the model requires that the utility function increases with income, due to the assumption that everyone chooses a point on the frontier of her budget set rather than in the interior (see below). We will not impose this assumption a priori but check whether it is satisfied by our unrestricted estimates. We will also see that the model does not require quasi-concavity of preferences, so that we do not have to impose this either. (This is because we use utility maximisation over a finite budget set, not requiring tangency conditions.) Thus we will not impose any a priori restrictions on the utility function in (1).

To allow the utility function to vary with taste shifters such as age and the number of children, the parameters $\alpha(p,q)$ may vary with a vector X of individual and family characteristics. In practice, it will be hard to disentangle effects of X via different $\alpha(p,q)$. In the estimates below therefore, only the parameter $\alpha(1,0)$—the coefficient of the linear term h in (1)—will be allowed to vary with $X: \alpha(1,0) = \beta_0 + X'\beta$. This makes it easy to interpret the results, since it implies that the marginal utility of leisure depends on X only through the additive term $X'\beta$. Thus the signs of the coefficients in β immediately determine the impact of the taste shifters on the marginal utility of leisure and on labour supply. There is no theoretical reason not to allow more of the $\alpha(p,q)$ to depend on X, or not to allow for a non-linear effect of X on these parameters. A specification that is fully non-parametric in X would obviously require this. The series expansion we use makes the utility function flexible in h and y but not in X. Practical limitations due to curse of dimensionality problems and limitations of the data prevented us from experimenting with specifications that are more flexible in X. Moreover, the utility function is only identified up to a monotonic transformation which may depend on X, implying that complete flexibility in terms of X would lead to an unidentified model (cf., e.g., Pollak and Wales, 1979).

Apart from heterogeneity through observed characteristics X, preferences may also vary with unobserved characteristics. To incorporate this, the Hausman type models typically add an additive random preference error term to one of the parameters of the utility function. We follow the same strategy and, in line with the way X is allowed to enter, assume that unobserved heterogeneity (u_{rp}) enters through the parameter $\alpha(1,0)$:

$$\alpha(1,0) = \beta_0 + X'\beta + u_{\mathrm{rp}}, \quad u_{\mathrm{rp}} \sim \mathrm{N}(0, \sigma_{\mathrm{rp}}^2). \tag{2}$$

The assumption that the distribution is normal is mainly made for convenience, and a more flexible distribution could in principle be used instead. For example, a mixture of normally distributed random variables could be used, or a discrete distribution with a finite number of mass points (cf. Heckman and Singer, 1984). These interesting non-parametric extensions will not be considered in our empirical work, since we already found it hard to accurately estimate the variance σ_{rp}^2 in the normal specification (2), with the standard error on the estimate of σ_{rp} exceeding its point estimate.

[1] The constant term $\alpha(0,0)$ will be normalised to 0.

2.2. Constraints

Labour supply is based upon utility maximisation under constraints. An obvious constraint is the budget restriction: different choices of the number of hours worked lead to different net family incomes. To determine net family income as a function of the wife's working hours, we need her gross earnings, other household income (husband's earnings, child benefits, asset income), taxes, and potential unemployment or social security benefits. The components of other household income are observed or can be computed directly. To determine the wife's earnings for each number of working hours, we assume that her gross hourly wage rate does not depend on her hours worked. This is a common assumption in most of the structural labour supply literature, although exceptions exist (see Moffitt, 1984, or Tummers and Woittiez, 1991). The assumption makes it possible to compute gross earnings for each number of working hours for those women who work, under the additional assumption that wage rates in the data do not suffer from measurement error.

For non-workers, we need to predict the before tax wage rate. For this purpose, and to be able to take account of measurement error in observed wage rates, we explicitly need to incorporate a wage equation. To account for selectivity of observed wages in a way consistent with the labour supply model, we estimate the wage equation jointly with the labour supply model.[2] The parameters in the wage equations are then used to predict the wages of non-workers. Because the labour supply model is non-linear in wages, wage rate prediction errors have to be taken into account to get consistent estimates of the labour supply model (the estimation technique is given below). The wage equation we use is given by

$$\begin{aligned}
&\mathrm{Log}\, w = Z'\gamma + u_w + e_w,\\
&u_w \sim \mathrm{N}\ (0, \sigma_u^2)\ \text{(unobserved heterogeneity in wages)},\\
&e_w \sim \mathrm{N}\ (0, \sigma_e^2)\ \text{(measurement error)},\\
&u_w,\ e_w \text{ independent of each other and of other}\\
&\quad \text{error terms in the model.}
\end{aligned} \tag{3}$$

Here w is the observed wage rate, which possibly contains measurement error e_w. According to simulation results of Blomquist (1996), such measurement error could substantially bias the estimates of the elasticities of interest, and this is the reason why we incorporate it. Note that σ_e and σ_u are separately identified because the wage entering the labour supply part of the model includes u_w but not e_w. Thus a more general interpretation of e_w might be the part of the wage rate that is job or hours specific, and that is not used in the labour supply decision.[3]

[2] Here we follow Gong and van Soest (2002) rather than Van Soest (1995). In the latter, the wage equation is estimated separately using a standard Heckman selection model (Heckman, 1979).

[3] The specification of the error term in the wage equation is different from the one in earlier studies. In Gong and van Soest (2002), a measurement error is not included explicitly, but the wage error is correlated to the random preference term. The correlation is interpreted as an indication of measurement error. Here we make this more explicit.

We do not allow for measurement errors on other variables. If women's wages are measured with error, men's wages probably contain error as well, and in our model this would lead to measurement error in other income. This could be included in a similar way, but would require an equation for other income. Results of Blomquist (1996) suggest that measurement errors in other income induce much less bias than measurement errors in the wage rate. For this reason and to keep things simple, we do not allow for measurement errors in other income.

In the sample we use for estimation, all husbands work, and they usually earn so much that social assistance benefits for the family do not apply: family income excluding the wife's own earnings typically exceeds the official minimum standard of living, which depends on age, marital status and family composition. For the few families in which husband's earnings are low, we incorporate social assistance benefits: if family income is below the official minimum standard of living, it is increased up to the minimum standard of living threshold. We do not model unemployment insurance benefits, because the temporary nature of these is incompatible with our static. Moreover, the data do not provide enough information to compute the unemployment benefits that non-workers would be entitled to. The most important benefits for our purposes are child benefits, which do not depend on earnings or labour market status.

Following Van Soest (1995), the budget constraint under which the individual maximises utility is approximated by a finite number of points. To make the model as close as possible to a continuous model, we will use a very fine grid size: we take working hours per week that are multiples of 10 minutes $(0, 1/6, 1/3, \ldots, 60)$.[4] This gives 361 points for each individual: (h_j, y_j), $j = 0, \ldots, 360$, where y_j is after tax family income if the wife works $h_j = j/6$ h per week. The model with 361 points will be used as the benchmark model for which we present all the results. We have also estimated the model using 16 points with intervals of 4 hours $(0, 4, \ldots, 60$ hours per week), and will briefly discuss these results as well.[5] The general conclusion is that there is hardly any difference between results and policy implications based upon models with 361 and 16 points, confirming the conclusions of Van Soest (1995) and Gong and van Soest (2002).

2.3. *Fixed costs of working*

Models without fixed costs of working, in which the utility function explains participation as well as hours worked, typically appear to under-predict the number of non-workers and over-predict the number of small part-time jobs. Including fixed costs of working is one way to repair this, since fixed costs of working make not working more attractive than working few hours per week. The level of the fixed costs may depend on individual and household characteristics Z. We model them as $\mathrm{FC} = Z'\delta$, where Z is a vector of individual and family characteristics.[6] In computing the values of the utility function $U(h_j, y_j)$, we then replace income y_j by $y_j - \mathrm{FC}$ if the wife

[4] Observed hours are truncated at 60.

[5] Observed hours worked per week are then rounded to multiples of four and truncated at 60.

[6] Unobserved heterogeneity in fixed costs can be allowed for by adding another error term. We experimented with this but it did not lead to significant improvement.

works, i.e., for $j>0$. If U is increasing with income, positive fixed costs decrease the utility of working compared to the utility of not working, and therefore decrease the probability of participation.

Fixed costs were also used by Callan and van Soest (1996) and Euwals and van Soest (1999). An alternative way to explain the lack of part-time jobs is given by Dickens and Lundberg (1993), Tummers and Woittiez (1991), and Van Soest et al. (1990), who model job offer probabilities for part-time jobs. Van Soest (1995) uses disutilities of part-time jobs, reflecting search costs of jobs with irregular hours. These methods attain the same goal as the fixed costs: they lead to a model that can reproduce the participation rate in the data as well as the sample mean of hours worked. The choice between the three seems a matter of taste; we chose fixed cost because it is economically plausible and hardly complicates the model or the estimation procedure.

As explained above, the intuitive explanation why fixed costs are identified is the lack of observations with a small positive number of working hours. While this argument is valid for a restrictive specification of the utility function that limits the way in which utility can vary with working hours, the argument would no longer hold if the specification of the utility function were fully non-parametric. For such a specification, the utility function itself could pick up the gap in the sample distribution of working hours, by assigning lower utility to small non-zero values of working hours that are sparse in the data. Thus it seems that the fixed costs are non-parametrically unidentified. In our specifications, the identification problem does not arise, due to the restrictive way in which the taste shifters enter the utility function and the fixed costs. Still, this seems a rather unnatural way to obtain identification. The way to avoid this would be not to include fixed costs explicitly in the non-parametric (higher polynomial expansion order) model, and to consider the utility function as an evaluation of preferences in which fixed costs are already captured. We have not done this since it hampers a fair comparison with the parametric (lower order) models.

2.4. Alternative specific error terms

The only error terms included so far are random preferences and errors in the wage equation. In addition, we introduce alternative specific error terms as follows:

$$u(h_j, y_j) = U(h_j, y_j) + \varepsilon_j, \quad j = 0, \ldots, 360. \tag{4}$$

We assume that the ε_j are independent and follow a standard (GEV I) extreme value distribution. We assume that the answer to the desired hours question is based upon maximising $u(h_j, y_j)$ rather than $U(h_j, y_j)$. The error ε_j can be interpreted as the random part of the evaluation of alternative j.

There are several reasons why these random errors are incorporated. First, they can be interpreted as unobserved job characteristics. Aaberge et al. (1999) describe an economic model leading to these errors. Second, they are needed to give non-zero probability to choices that cannot be optimal for any value of the random preference term. Such choices may very well exist in case of a non-convex or discontinuous budget set, where some points on the budget frontier may give low family income compared to adjacent points. In this sense, they play the same role as the optimisation or measurement errors in the traditional Hausman (1985) model. Third, it is attractive to include the ε_j from a computational point of view, since they facilitate simulated maximum likelihood estimation by smoothing the approximation of the likelihood, and can thus be seen as a smoothing device. Keane and Moffitt (1998) use the same type of error terms, but, enforcing the third interpretation, impose that the ε_j have a small variance compared to the variance in $U(h_j, y_j)$. We do not make such an assumption and thus allow for all interpretations.

Due to the assumption on the distribution of the ε_j, the model is similar to a multinomial logit model. The probability that an individual chooses alternative j, conditional on wage rates, tax and benefit rules, exogenous variables, and random preferences, is given by

$$P[j] = \exp\{U(h_j, y_j)\} \bigg/ \sum_k \exp\{U(h_k, y_k)\}. \tag{5}$$

Here the summation in the denominator is over all points in the choice set.

$P[j]$ increases with $U(h_j, y_j)$ (given the other $U(h_k, y_k)$). If U increases in income, the utility of working increases with the (before and after tax) wage rate. The utility of not working is not affected by the wage rate. Thus the participation probability increases with the wage. This illustrates that the participation decision is fully incorporated in the structural model, other than in many studies considering labour supply conditional on participation.

The assumption that the ε_j are independent and follow an extreme value distribution seems restrictive and introduces a parametric feature into the model. More flexible assumptions such as multinomial probit are possible, in principle. See the discussion in Hajivassiliou and Ruud (1994). There are two reasons why we retain the GEV assumption in the current paper. First, as already mentioned, the estimates of the variance of the random preference term are very imprecise, suggesting that pinning down the structure of the error terms is difficult with the data at hand. Moreover, McFadden and Train (2000) show that any probability structure of the discrete choices can be captured using the GEV I errors in combination with a non-parametric specification of unobserved preference heterogeneity. Thus the GEV I assumption by itself is not restrictive. Second, as shown below, the similarity with the multinomial logit model makes it possible to relate the estimation problem to maximum entropy estimation and to define an information index.

2.5. Coherency

An important issue in the piecewise budget constraint model is coherency. If preferences are not quasi-concave in some relevant region, the model may not have a well-defined unique solution for a non-zero probability set of values of the error terms. This implies that the probabilities used in the likelihood function do not add up to one, and that maximising the likelihood can lead to inconsistent parameter estimates. Van Soest et al. (1993) give an example where the latter is indeed the case. They argue that coherency should be imposed a priori before estimation. MaCurdy et al. (1990) show that imposing coherency in the linear labour supply model implies that labour supply cannot be backward bending. Thus imposing coherency conditions limits the flexibility of the specification of preferences. Van Soest et al. (1993) show that quasi-concavity of preferences is sufficient but not necessary to guarantee coherency, and confirm the result that imposing coherency in a restrictive specification leads to bounds on the potential elasticities and policy effects. Blomquist (1995) shows that this problem is not unique to ML-estimation but also plays a role if another estimation technique (instrumental variables) is used.

In the current set up, there are two reasons why the coherency problem does not arise and conditions limiting flexibility can be avoided. First, the coherency problem in the Hausman (1985) model is due to solving the utility maximisation problem using Kuhn–Tucker first order conditions. If preferences are not quasi-concave, there may be multiple solutions to these first order conditions. The multiple-regime econometric model based upon the Kuhn–Tucker conditions may then have zero or more than one solution. Our set up does not rely on tangency conditions or duality theory, since utility is maximised over a finite set. Since the ε_j have a continuous distribution, the probability that two points have optimal utility is zero, and the model has a unique solution with probability one and is coherent, irrespective of the shape of the utility function.

A second danger exists, however. An implication of the MaCurdy et al. (1990) critique is that a seemingly flexible functional form may not be flexible anymore once quasi-concavity or monotonicity is imposed. Even though we do not impose these conditions explicitly, it might still be the case that the structure of the model implicitly will enforce the estimates to satisfy quasi-concavity. For example, a wrongly shaped utility function would lead to high probabilities of choosing the corners of the budget frontier (0 or TE hours of work), and maximum likelihood estimates will avoid this shape if observations at these corners are sparse. Thus even if coherency is not imposed, the question whether quasi-concavity combined with functional form does not limit the range of elasticity values or policy effects remains relevant. In our case, however, we use a flexible functional form of the utility function. Even if quasi-concavity of preferences on some relevant region of (h, y) space were imposed, we would not impose more than that, because our series expansions approximate any quasi-concave utility function arbitrarily closely. Thus the problem in MaCurdy et al. (1990) that due to a restrictive functional form imposing coherency or quasi-concavity immediately bounds the range of possible elasticities, will not occur.

We will not impose quasi-concavity but we will check ex post whether estimated preferences are quasi-concave. While this is not necessary for the interpretation of the model (only utility in the finite choice sets matters), it would help to reconcile our findings with those of the Hausman approach. For economic interpretation (and meaningful policy simulations), we need that utility increases with income. We will check this ex post without imposing it a priori.

2.6. Estimation

Due to the multinomial logit nature of the model, estimation by maximum likelihood would be straightforward if random preference terms were observed and all wages were observed without measurement error. In that case, the likelihood would follow directly from (2), (3) and (5), since the $U(h_j, y_j)$ would be known functions of parameters, explanatory variables, the observed wage rate, and the known random preference term. The likelihood contribution of a given individual would be her wage density (following from (3)) multiplied by the probability in (5). But we do not observe the random preference term and may observe wages with error or not at all. As a consequence, the likelihood contribution of a given observation is given by the mean of the appropriate expression according to (2), (3) and (5), with the mean taken over the unobserved errors. This mean is a two-dimensional integral. Such an integral can be approximated by conventional numerical (quadrature) routines. A convenient alternative which also works for dimensions higher than two, is simulated maximum likelihood: the integral is replaced by a simulated average based upon R independent draws from the (multivariate normal) distribution of the unobserved errors, conditional upon the observed wage rate. Due to the law of large numbers, the approximation will be accurate if R becomes large. With independent draws across observations, it can be shown that the approximation is accurate enough to make simulated maximum likelihood asymptotically equivalent to exact maximum likelihood if R tends to infinity faster than the square root of the number of observations (see, for example, Hajivassiliou and Ruud, 1994). We will use $R=20$. The sensitivity of the results for the choice of R is analysed by Gong and van Soest (2002), who find that $R=20$ is large enough. We will obtain the same conclusion in our Monte Carlo study.

The ε_j make estimation easier. Without them, the likelihood contribution conditional on the unobserved error terms would be either 0 or 1. The simulated likelihood would not be continuous in the parameters, locating the maximum would be harder, and zero contributions would arise. Adding the ε_j smoothes the likelihood and bounds it away from zero. Thus adding the ε_j can be seen as a smoothing device. This is the interpretation of Keane and Moffitt (1998), who fix the variance of ε_j at a small value, and also impose a normalisation on the systematic part of the utility function. This a priori limits the share of the variance of the ε_j in the total variance of $u(h_j, y_j)$. We normalise the variance of ε_j only, and do not impose an additional scale normalisation on the utility function.

2.7. The link to information theory

The similarity of the model to a standard multinomial logit model makes it possible to construct an information index based upon the likelihood. To make the paper self-contained, we first discuss how this works in the context of a standard multinomial logit model, following Soofi (1992) and Golan et al. (1996, Section 3.2).

Consider a multinomial choice problem where N respondents $i=1,\ldots,N$ choose from J alternatives $j=1,\ldots,J$. The observed choices in the data are denoted by y_{ij}, where $y_{ij}=1$ if respondent i chooses alternative j and $y_{ij}=0$ otherwise ($i=1,\ldots,N$; $j=1,\ldots J$). The choice probabilities depend on a vector x_{ij} of respondent and alternative specific characteristics. Let p_{ij} be the probability that respondent i chooses alternative j. The entropy measure is the following function of the matrix p of all p_{ij}.

$$H(p) = -1/N \sum_{i=1,\ldots,N} \sum_{j=1,\ldots,J} p_{ij} \log p_{ij}. \tag{6}$$

The maximum entropy (ME) estimator for p is defined as the value of p that maximises $H(p)$, subject to the information-moment constraints

$$\sum_{i=1,\dots,N} x_{ij}(y_{ij}-p_{ij})=0, \quad j=1,\dots,J, \tag{7}$$

as well as N normalisation conditions

$$\sum_{j=1,\dots,J} p_{ij}=1, \quad i=1,\dots,N. \tag{8}$$

Solving this maximisation problem is straightforward. As is well-known, the solution is given by the multinomial logit (MNL) probabilities

$$p^*_{ij}=\exp\left\{\beta^{*\prime}x_{ij}\right\}\Bigg/\sum_{k=1,\dots,J}\exp\left\{\beta^{*\prime}x_{ik}\right\}. \tag{9}$$

Here β^* is the negative of the vector of the Lagrange multipliers associated with (7), as well as the vector of ML estimates of β in the MNL model with probabilities

$$\mathrm{P}\{y_{ij}=1|x_{i1},\dots x_{iJ};\beta\}=\exp\left\{\beta'x_{ij}\right\}\Bigg/\sum_{k=1,\dots,J}\exp\left\{\beta'x_{ik}\right\}. \tag{10}$$

If the multinomial logit model is a correct specification of the data generating process, then for large N, the maximum entropy value $H(p^*)$ is approximately equal to the negative of the average log likelihood value:

$$\begin{aligned}&-1/N\sum_{i=1,\dots,N}\sum_{j=1,\dots,J} y_{ij}\log\mathrm{P}\{y_{ij}=1|x_{i1},\dots x_{iJ},\beta^*\}\\&\quad\approx -1/N\sum_{i=1,\dots,N} p^*_{ij}\log p^*_{ij}=H(p^*).\end{aligned}$$

This helps to motivate the use of several goodness of fit measures based on ME information diagnostics. Golan (1988) introduced the normalised entropy $H(p^*)/H(p^0)$, where p^0 corresponds to the uniform distribution with probabilities $p^0_{ij}=1/J$. $H(p^0)$ is the maximum entropy that can be attained if only the normalization constraints in (8) are imposed and not the information-moment constraints in (7). The non-negative difference $H(p^0)-H(p^*)$ can be seen as the uncertainty reduction due to the data. Soofi (1992) introduced the information index

$$I(p^*)=1-H(p^*)/H(p^0). \tag{11}$$

If the data are not very informative in the sense that the x_{ij} have little explanatory power, p^* will be close to p^0 and $I(p^*)$ will be close to zero. On the other hand, if the data are informative and predictions are almost perfect, $H(p^*)$ will be close to zero and the information index will tend to 1. Thus the interpretation of $I(p^*)$ is similar to the R^2 goodness of fit measure in a linear regression model. In terms of average sample log-likelihoods log $L(\beta)$, we get

$$H(p^*) \approx -1/N \sum_i \log L_i(\beta^*) \text{ (}-\log \text{likelihood of the MNL model)}$$

$$H(p^0) \approx -1/N \sum_i \log L_i(0) \text{ (}-\log \text{likelihood if } \beta = 0,$$

$$\text{such that } p_{ij} = p_{ij}^0 = 1/J)$$

and thus $2NH(p^0)I(p^*) = 2N[H(p^0) - H(p^*)]$ is approximately equal to the usual test statistic of a likelihood ratio test, and can be used for testing purposes in an ME framework.

The choice model developed in the current paper cannot be written in the standard MNL form (10), for three reasons: unobserved wage rates of non-workers, measurement errors in observed wage rates, and unobserved heterogeneity in preferences. Let us ignore the fact that the wage equation is estimated jointly with the labour supply part of the model, and consider the labour supply part of the model only. If we denote the unobserved components of both preferences and the wage rates (excluding the measurement errors, which do not enter the choice part of the model) by a vector ς, then the likelihood contribution of respondent i choosing alternative j can be written as

$$\mathrm{P}\{y_{ij} = 1 | x_{i1}, \ldots, x_{iJ}\} = \mathrm{E}[\mathrm{P}\{y_{ij} = 1 | x_{i1}, \ldots, x_{iJ}, \varsigma\}]. \tag{12}$$

Here $\mathrm{P}\{y_{ij}=1|x_{i1},\ldots,x_{iJ},\varsigma\}$ has the MNL form given in (10). The expectation is taken over ς.

We need to generalise the diagnostic $I(p^*)$ to account for the unobserved heterogeneity (the error terms ς). Using concavity of the log, we get, for each observation i:

$$\begin{aligned} 0 &\geqslant \log \mathrm{P}\{y_{ij} = 1 | x_{i1}, \ldots, x_{iJ}\} = \log \mathrm{E}[\mathrm{P}\{y_{ij} = 1 | x_{i1}, \ldots, x_{iJ}, \varsigma\}] \\ &\geqslant \mathrm{E}[\log \mathrm{P}\{y_{ij} = 1 | x_{i1}, \ldots, x_{iJ}, \varsigma\}]. \end{aligned}$$

Let θ denote the parameters of the labour supply model (including the variance of the random preference term). Then this gives

$$\begin{aligned} 0 &\geqslant \mathrm{Max}_\theta 1/N \sum_i \log \mathrm{E}[\mathrm{P}\{y_{ij} = 1 | x_{i1}, \ldots, x_{iJ}, \varsigma\}] \\ &\geqslant \mathrm{Max}_\theta 1/N \sum_i \mathrm{E}[\log \mathrm{P}\{y_{ij} = 1 | x_{i1}, \ldots, x_{iJ}, \varsigma\}] \\ &\geqslant 1/N \sum_i \log 1/J = \log 1/J. \end{aligned}$$

The law of large numbers implies that, for a large random sample, we can delete the expectation in the final term and replace it by one random draw ς_i per observation, so that, for large N:

$$0 \leqslant -\mathrm{Max}_\theta 1/N \sum_i \log L_i(\theta) \leqslant -\mathrm{Max}_\theta 1/N \sum_i \log L_i(\theta, \varsigma_i\} \leqslant \log J.$$

This can be interpreted in a similar way as in the standard MNL model. The first term, 0, is the entropy of the data at the limit. The third term approximates the entropy of multinomial choice data where a random draw of the vector of unobserved components ς_i is added to each observation. This corresponds to $H(p^*)$ in a standard multinomial logit model. The second term is minus the average log likelihood in the data. The difference between the second and third term is due to the presence of unobserved heterogeneity.

Instead of comparing various models (with different specifications of preferences, and thus with different degrees of flexibility) in terms of their likelihood, we will also compare them on the basis of the goodness of fit measures derived from the inequalities above:

$$I_1 = 1 + \text{Max}_\theta 1/N \sum_i \log L_i(\theta)/\log J = 1 + 1/N \sum_i \log L_i(\theta^*)/\log J,$$

$$I_2 = 1 + \text{Max}_\theta 1/N \sum_i \log L_i(\theta, \varsigma_i)/\log J.$$

I_1 is the information index directly based upon the log likelihood. I_2 is the information index for an MNL likelihood with random draws of unobserved heterogeneity. In a large sample and if the model is correctly specified, we expect that I_1 exceeds I_2. The difference can be interpreted as the gain in uncertainty reduction due to enriching the model with unobserved heterogeneity.

3. Data and estimation results

The data are drawn from May 1995 wave of the Dutch Socio-Economic Panel (SEP). This is a panel consisting of about 5,000 households. It is representative for the Dutch population excluding people living in nursing homes, etc. We focus on married or co-habiting women in the age group 16–64 whose partners have a paid job with observed earnings. We exclude women who are full-time students, receive full-time disability benefits, or receive pensions or other retirement benefits. This leads to a sample of 1,794 women; 1,100 of these have a paid job.

We aimed at constructing the dependent labour supply variable in such a way that demand side restrictions do not play a role. Thus we wanted to use desired working hours (per week) instead of actual hours worked. For people who are looking for a(nother) job, the survey has information on how many hours they would like to work per week in this (new) job. We consider this number rather than actual hours as their desired labour supply. For those who are not looking for a(nother) job, however, we have no further information, and have to assume that desired hours are given by actual hours. Of the 694 non-workers in the sample, 116 are looking for a job, and thus have positive desired hours. Of the 1,100 workers in the sample, 152 are looking for another job with desired hours different from actual hours.

Earnings in the SEP are measured as gross earnings in the year 1994, retrieved from the respondent's tax files. These earnings can only be used to compute an hourly wage rate for the job held at the time of the survey in May 1995 for people who have not changed jobs in 1994 or between January 1995 and May 1995. For those who changed jobs in that period, earnings are set to missing. This concerns 55 observations. Table 4 in the appendix contains the definitions of the variables we use in the analysis and some sample statistics. Non-participation among married women in the Netherlands is rather large, although it has fallen substantially during the past two decades. In our sample, which excludes those who are not available for work such as students or disabled, the net participation rate (excluding those who are looking for work) is 60.9%, while the gross participation rate (including job searchers) is 67.3%.

We have estimated the benchmark model for $K=1$ to 5. In Table 5 (see Appendix) we present the results for $K=5$. Most parameters in the utility function cannot be interpreted directly. The exceptions are the interactions between hours worked and characteristics, since these parameters determine how the marginal utility of leisure changes with characteristics. These results show that age is insignificant, while the presence of children increases the marginal utility of leisure (i.e., reduces the marginal utility of hours worked) and thus reduces labour supply. This effect is still stronger when there are children of pre-school age. The findings on the impact of children are in line with the bulk of the labour supply literature (see Nakamura and Nakamura, 1990, for example). Random preferences play a minor role: σ_{rp} is small and the standard error exceeds the estimate.

We find that children reduce fixed costs, which seems counterintuitive. Combined with the effect of children on preferences, this finding could mean that for women with children, working a small number of hours per week is attractive. Fixed costs fall with age until approximately age 46.

The wage equation estimates correspond to common findings in the human capital literature, with an increasing age pattern until about age 41 and higher wages for the higher educated. The estimates of the standard deviations σ_u and σ_e suggest that the labour supply decision is largely based on the predicted wage. Most of the unpredicted part of the wage is not used in the labour supply decision, for example, because this is measurement error.

All the results discussed above are similar for the second order and fifth order models. They are also very similar for the second, third and fourth order models. Only the first order estimates give a different picture, with, for example, an even larger role of the measurement error e_w compared to u_w. The results above also remain unchanged if 16 4-hours intervals are used instead of 361 10-minutes intervals.

To choose between the models of different orders, Table 1 presents their likelihoods, their AIC values, and the two information indexes I_1 and I_2 introduced in Section 2. According to all these criteria, the first order model is clearly outperformed by all other models. Differences among the higher order models are much smaller. According to likelihood ratio tests, the second order model is not rejected by the third order model at the 5% significance level, but all lower order models are rejected against the most general model, the fifth order model. The AIC values lead to the same conclusion: the fifth order model outperforms the other models. The information index numbers again show that the third order model hardly improves the goodness of fit compared to the second order model, but that the fourth and fifth order models lead to somewhat larger improvements. The largest difference is between first and second

Table 1
Log likelihoods, AICs and information indexes (real data)

Specification	# Parameters	Likelihood	AIC	I_1	I_2
1st order	22	−8438.79	9.432	0.2620	0.2621
2nd order	25	−8088.92	9.046	0.2952	0.2925
3rd order	29	−8084.89	9.051	0.2956	0.2925
4th order	34	−8073.74	9.039	0.2966	0.2934
5th order	40	−8044.71	9.013	0.2994	0.2959

order models, however. There is hardly any difference between the values of I_1 and I_2, suggesting that enriching the model with heterogeneity hardly reduces uncertainty. This is in line with the small role of random preferences and the fact that most of the unobserved variation of wages is attributed to measurement error and does not play a role in the choice part. For the higher order models, I_1 is larger than I_2, in line with our expectations based on large samples. That this is not so for the first order model can again be seen as an indication that the first order model is seriously misspecified.

In the sequel, we will focus on the first, second and fifth order model, since all results for the third and fourth order models are in between the results for the second and fifth order model.

Fig. 1 shows some estimated indifference curves for the first, second and fifth order models. Utility is increasing with income in almost all data points, although we did not impose this a priori. Utility is also usually increasing with leisure, except at low hours for the fifth order model.

The indifference curves for the first order model are by definition linear, so quasi-concavity is not an issue there. The quadratic indifference curves for the second order model imply quasi-concavity of preferences for all values (h, y). For the fifth order model, the indifference curves are convex in most of the relevant range of (h, y)-space, but not everywhere. In particular, quasi-concavity of preferences is violated at points in (h, y) space with high values of working hours.

Fig. 2 presents labour supply curves, constructed in a similar way as traditional labour supply curves. Hours worked is drawn as a function of the woman's own wage rate. The curves concern a benchmark individual (age $= 40$, non-female income $= 1,000$, 1 young child, and random preference term set to zero). We show expected hours worked in case of a linear budget constraint (no fixed costs, no taxes or benefits), computed as the probability weighted mean of the possible hours values $0, 1/6, 2/6, \ldots, 60$ in the model. The probabilities are computed from the model estimates (using (5)). Again, the first order model gives very different outcomes from the other models. The other models lead to labour supply curves that are similar to each other in the range of low wage rates, but show some larger differences for high wage rates. For the second, third and fourth order model, labour supply is everywhere forward bending. For the fifth order model, labour supply is forward bending except at very high wage rate levels.

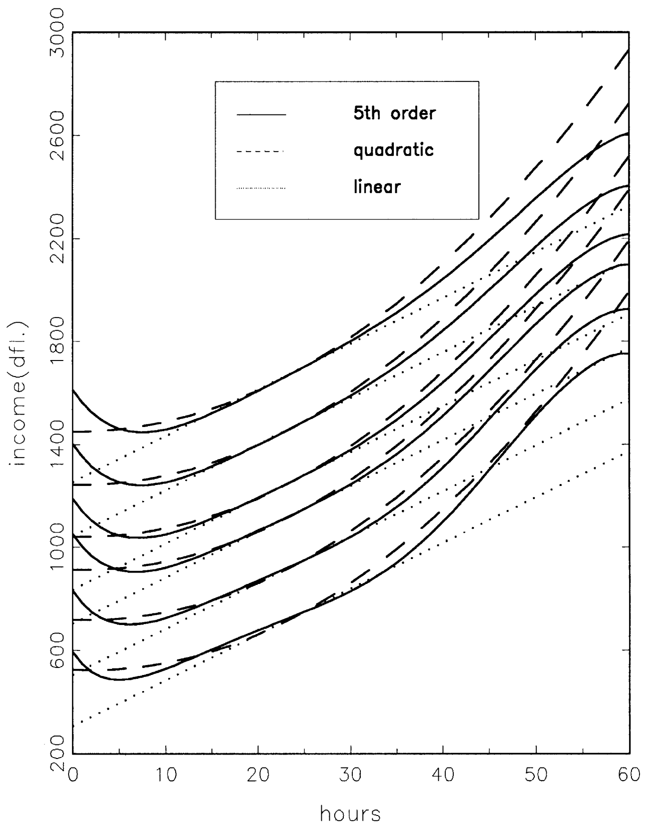

Fig. 1. Indifference curves benchmark individual.

3.1. Elasticities

Fig. 2 shows the sensitivity of labour supply to the wage rate for a benchmark individual. Since elasticities vary over the sample and the model is very non-linear, elasticities for the benchmark individual may not say much about aggregate elasticities. We define the aggregate (own) wage elasticity of labour supply of some given group of women as the percentage change in total desired hours of that group if all their before tax wage rates rise by 1%. Unlike some other elasticities used in the literature, this definition takes full account of the impact of the wage rate on the participation decision (with desired hours equal to zero for non-participants). We also show which share of the elasticities is due to the effect on participation.[7] Our elasticities are based on increasing gross wage rates and leaving the tax system unaffected. Thus they correspond to what Blomquist (1996) calls the Mongrel labour supply function rather than the

[7] We look at desired hours and not at actual hours. Participation is defined as having positive desired hours.

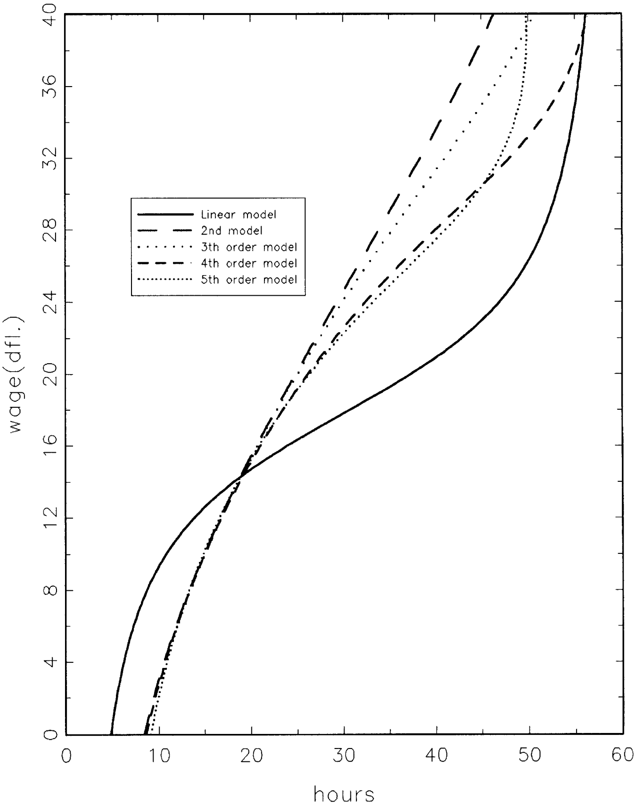

Fig. 2. Labour supply curves benchmark individual.

labour supply curves in Fig. 2. Since the benchmark for policy analysis will be the actual tax system and not some linear system, this definition considers wages in the relevant range. On the other hand, it has the drawback that the way in which net wage rates change is endogenous. On average, they will change by slightly less than 1%, due to the progressive nature of the tax rules.

The elasticities can be computed from model simulations, first using actual (predicted) wage rates, and then with all wage rates increased by 1%. Standard errors are computed by deterministic bootstrapping, repeating the calculations for new parameter values of the model drawn from the estimated distribution of the parameter estimates. In a very similar way, we have also calculated the elasticity and the sensitivity of participation with respect to the husband's wage rate. Since the husband's earnings reflect the bulk of family income other than the woman's own earnings, these are approximately the same as other income elasticities.

The first four columns of Table 2 contain the results for each of the five models, for all women and separately for the low and high educated. In line with Fig. 2, the elasticities according to the first order model deviate substantially from those according

Table 2
Changes in participation, elasticities of hours worked, and policy effects (real data, bootstrapped standard errors are in parentheses)

Model	Husband's wage		Own wage		Policy effects	
	Change in participation (in %-points)	Elasticity of hrs worked (in %)	Change in participation (in %-points)	Elasticity of hrs worked (in %)	Change in participation (in %-points)	Change in hrs worked (in %)
5th order						
All	−0.107	−0.173	0.483	1.155	−3.39	4.31
	(0.025)	(0.046)	(0.050)	(0.109)	(0.49)	(0.70)
Low educated	−0.119	−0.215	0.507	1.232	−4.14	4.08
	(0.027)	(0.047)	(0.054)	(0.123)	(0.57)	(0.76)
High educated	−0.061	−0.058	0.383	0.945	−0.34	4.93
	(0.020)	(0.067)	(0.042)	(0.094)	(0.27)	(0.84)
4th order						
All	−0.101	−0.163	0.482	1.143	−3.20	4.43
	(0.028)	(0.046)	(0.047)	(0.11)	(0.43)	(0.63)
Low educated	−0.114	−0.203	0.507	1.226	−3.94	4.25
	(0.031)	(0.050)	(0.052)	(0.13)	(0.51)	(0.70)
High educated	−0.050	−0.053	0.380	0.915	−0.17	4.92
	(0.019)	(0.048)	(0.037)	(0.081)	(0.22)	(0.61)
3rd order						
All	−0.109	−0.173	0.453	1.067	−3.01	4.08
	(0.028)	(0.040)	(0.044)	(0.10)	(0.47)	(0.53)
Low educated	−0.120	−0.205	0.471	1.145	−3.71	3.96
	(0.030)	(0.046)	(0.047)	(0.11)	(0.55)	(0.58)
High educated	−0.065	−0.083	0.376	0.850	−0.15	4.42
	(0.018)	(0.032)	(0.038)	(0.076)	(0.25)	(0.50)
2nd order						
All	−0.100	−0.152	0.439	1.043	−2.74	4.35
	(0.021)	(0.035)	(0.045)	(0.105)	(0.35)	(0.55)
Low educated	−0.108	−0.177	0.455	1.121	−3.38	4.35
	(0.023)	(0.038)	(0.048)	(0.119)	(0.40)	(0.61)
High educated	−0.067	−0.083	0.371	0.826	−0.10	4.37
	(0.015)	(0.028)	(0.036)	(0.071)	(0.23)	(0.45)
1st order						
All	−0.038	−0.090	0.505	1.971	−2.20	14.26
	(0.0043)	(0.010)	(0.056)	(0.21)	(0.36)	(1.57)
Low educated	−0.041	−0.109	0.515	2.140	−2.86	15.95
	(0.0048)	(0.013)	(0.059)	(0.25)	(0.42)	(1.86)
High educated	−0.025	−0.043	0.465	1.554	0.53	10.08
	(0.0029)	(0.0053)	(0.045)	(0.14)	(0.25)	(1.04)

to the other models. The four higher order models, however, lead to similar elasticities, with overlapping confidence intervals. The own wage elasticity is somewhat above one. While this is not out of line with other findings for the Netherlands (cf. Theeuwes, 1988, for example) or other countries (Killingsworth and Heckman, 1986; Blundell and MaCurdy, 1999), it is somewhat higher than recent findings with similar models (cf. Van Soest and Das, 2001, and Vlasblom, 1998). The main reason seems to be that we have allowed for measurement error in the wage rates: if we set σ_e to zero, the own wage elasticity is about half as large. This difference is in line with the standard argument that measurement error biases the coefficients to zero if not properly accounted for. The cross-wage elasticity is about -0.17, in line with earlier findings for the other income elasticity (Theeuwes, 1988). The effects on participation presented in Table 2 are the changes in percentage points if husband's or own wage rates increase by 1%.[8] More than half of the cross-wage elasticity, and somewhat less than half of the own wage elasticity are due to an effect on participation.

Labour supply of low educated women is of particular interest from a policy point of view since their participation rates are lower than for other women. We find that labour supply of the low educated is somewhat more sensitive for both wage rates than labour supply of the high educated. Again, the various higher order models lead to the same conclusions here. All these conclusions remain virtually unchanged if a model with 4 hours intervals is used. For example, the own wage elasticity of hours worked then changes from 1.155 (std. error 0.11) to 1.259 (std. error 0.13).

3.2. Tax reform

To illustrate the usefulness of the fully structural model, we have analysed the potential consequences of a recently proposed tax reform, in which the non-convexities in the budget set close to zero hours of work play a large role. Such a reform can therefore not be analysed using the convex budget constraint model of Blomquist and Newey (1997), or other models not considering the participation decision or not allowing for non-convex budget sets.

We briefly describe the tax system and the proposed reform. More details are given in Van Soest and Das (2001). The budget constraints for a benchmark value of the husband's income (working full time with earnings equal to three times the minimum wage) are presented in Fig. 3.[9] The current system has individual taxation of the two spouses, but with one joint feature: the length of the first tax bracket (with a rate of 0%) depends upon earnings of the spouse. If both spouses work and both earn more than Dfl 8,600, then the tax-free allowance for both is Dfl 8,600. If the wife has no own income, the husband's tax-free allowance is Dfl 16,800—the wife's tax free allowance is transferred. If the husband earns more than Dfl 8,600, but the wife earns less than Dfl 8,600, the wife will transfer her allowance to the husband, so that her

[8] The elasticities of participation can be obtained by dividing these numbers by the predicted participation rates, approximately 0.673 for the whole sample, 0.635 for the low educated, and 0.832 for the high educated.

[9] We work with weekly hours and weekly income, while the tax system uses annual income. Thus weekly before tax income is multiplied by 365/7, the tax system is applied, and annual after tax income is divided by 365/7.

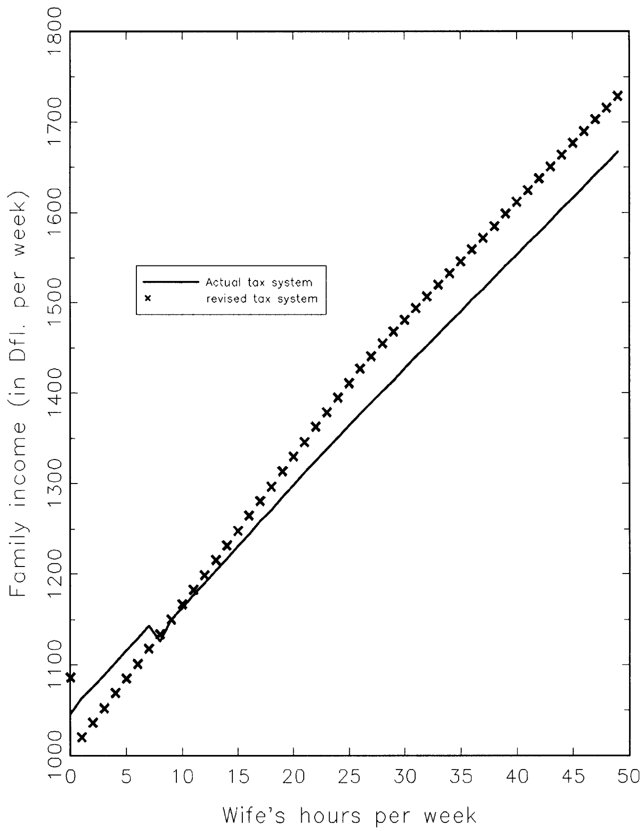

Fig. 3. After tax family income as a function of the hours worked by the wife. The husband earns three times the minimum wage and the wife's hourly wage is equal to 1.5 times the minimum (hourly) wage.

own allowance becomes Dfl 400 and her husband's allowance will be Dfl 16,800. The transfer possibility creates a disincentive for the woman to earn more than Dfl 8,600 if the husband's earnings are high. This is shown by the solid curve in Fig. 3, which gives net family income as a function of the wife's hours of work. There is a dip when the wife's earnings attain the maximum transfer threshold.

In a recent report (Ministry of Finance, 1997), the main ideas are sketched for a complete reform of many features of the tax system of the Netherlands. The main reforms concern increasing taxes on polluting activities, changing some of the VAT rates, and reducing taxes on labour. The latter implies changing the income tax system for private households. The report contains 21 specific proposals. We only look at the most radical type of reforms, which involves abolishing tax-free allowances for two earner families. [10] Only in genuine one-earner families, the tax-free allowance of the only earner would be increased. Extra tax revenues are used to lower the marginal

[10] We consider the basic version of option 3 in Ministry of Finance (1997). The proposal that has gone to parliament (Ministry of Finance, 1999) avoids the problem with small part-time jobs addressed below.

rates, so that the revision as a whole (also accounting for changes in other taxes) is revenue neutral.

The budget constraint for the benchmark family after the reform is given by the dotted line in Fig. 3. There is a discontinuity at 0 hours of work: as soon as the woman starts working, the additional tax-free allowance of the husband is lost, and family income falls. Thus the proposed reform creates a disincentive for women to accept a small part-time job.

The estimated effects of the policy reform are presented in the final columns of Table 2. Again, all higher order models give similar results, but the first order model does not. Since many women who now have a small part-time job would decide to stop working after the reform, participation would fall. This effect is concentrated in the group of low educated women. On the other hand, net earnings of full-time workers increase, and labour supply is stimulated in the sense that there is a substantial increase in hours worked, in spite of the negative effect on participation. Average hours (including zeros) would increase by between 4% and 5%. In the model with the 4 hours intervals, the effects are similar but somewhat smaller in absolute value.

4. Monte Carlo simulations

The goal of this section is threefold. First, we want to establish the finite sample properties of the estimator if the model is correctly specified. There are two reasons why there could be a problem. First, the model is non-linear and fairly complicated, and the finite sample bias of maximum likelihood in this type of model is an open issue. Second, we do not use exact maximum likelihood but simulated maximum likelihood, approximating the likelihood using $R=20$ draws for each observation (with independent draws across observations). While there is some evidence that the results are not very sensitive for the chosen value of R, this remains a potential source of bias for every new version of the model or every new data set.

The second purpose of the Monte Carlo simulations is to analyse the consequences of using a too restrictive model for the estimates of elasticities and policy effects. In the previous section, we have seen that the fifth order model is preferred to any of the lower order models using the information index, the AIC criterion or likelihood ratio tests. Still we did not find much difference between elasticities according to the various higher order models, suggesting that a second order model would be sufficient to get reasonable estimates. We want to confirm this in a situation where we know that the order of the series expansion is the only source of misspecification.

Third, the issue of measurement error in hours worked has until now not been addressed. In the original Hausman model, measurement error is explicitly incorporated, but in our model it is not. We want to check whether the presence of measurement error in hours worked can bias the results for the parameters of interest in our model.

We used the exogenous variables in the sample and the estimates of the fifth order model in Table 5 (see Appendix), to generate 100 new data sets. Using Eq. (3), wage rates (including measurement errors) are generated for all observations. Using Eq. (5) (and errors drawn from the GEV I distribution), optimal hours are then generated (i.e., the utility maximising element in the set $\{0, 1/6, 2/6, \dots, 60\}$ is determined). Wage rates of non-workers are then set to missing. The 100 new data sets are generated independently (conditional on the exogenous variables), and each of them consists of 1,794 (independent) observations.[11]

For each of the 100 new data sets, we estimated the fifth order and the second order model. (Measurement errors on hours worked will be discussed below.) Each of these estimations required numerical optimisation of a simulated likelihood function, and thus requires substantial computational effort. This is why we have restricted the Monte Carlo study to 100 new data sets, which is rather low compared to other Monte Carlo studies.[12] We believe, however, that the results are stable enough to be confident that a larger number would not lead to very different results.[13] Based upon each set of estimates, we repeated the calculations of elasticities and policy effects as presented in Table 2. In Table 3, we summarise these results. We focus on the sample as a whole, and do not distinguish between education levels. For each elasticity or policy effect, we present the mean and the "sample" standard deviation over the 100 new data sets.

The first two columns of Table 3 refer to the second and fifth order model without measurement errors in hours worked. The fifth order model can be used to answer the first question: is there a serious bias in the simulated ML estimates, either due to the small sample or due to the small value of R? We know that the true elasticities and policy effects are their point estimates for the fifth order model in the first row of Table 2. The result is striking: the numbers in the second column of Table 3 are quite close to those in the first row of Table 2. There is no evidence of any systematic bias in the point estimates of the elasticities. There is a larger difference in the estimates of the policy effect. Why the policy effects are harder to estimate than the elasticities is not so clear; perhaps it is because the former is a more complicated function of the parameters of the model than the latter.

The standard deviations in the second column of Table 3 can be compared to the estimated standard errors in the first row of Table 2. Standard errors in Table 2 are derived from the estimated asymptotic distribution of the (simulated) maximum likelihood estimates of the parameters in the model. They could be inaccurate due to finite sample bias, due to choosing R too small,[14] or due to model misspecification.[15] Comparing them to the standard deviations in Table 3 suggests that the standard errors in Table 2 are somewhat underestimated. Again however, the differences for the elasticities are quite small. For the policy effect on hours worked, the difference is somewhat larger,

[11] We do not generate actual hours, involuntary unemployment, or missing wage rates for workers. Thus, other than in the real data, for all those in the new data sets who want to work, the wage rate is observed. As a consequence, the Monte Carlo simulations may somewhat overestimate the accuracy of results that can be obtained with the real data.

[12] Blomquist (1996), for example, uses 400 replications. He has much simpler models (with fewer parameters), and only 602 observations, however.

[13] We also experimented with 75 new data sets, and obtained very similar conclusions.

[14] We have estimated the standard errors using the common ML technique, implicitly assuming that R tends to infinity at least as fast as the square root of the number of observations. This is the condition under which ML and simulated ML are asymptotically equivalent (Hajivassiliou and Ruud, 1994, p. 2419).

[15] We have used the outer product estimator of the information matrix; more robustness to misspecification might be obtained by also using the Hessian.

Table 3
Results of a Monte Carlo study without and with measurement errors in hours worked*

	Without measurement error		With measurement error			
			S.D. = 2		S.D. = 4	
	Estimated model		Estimated model		Estimated model	
	2nd order	5th order	2nd order	5th order	2nd order	5th order
Wage effects						
Husband's wage						
Change in participation	−0.104	−0.108	−0.104	−0.108	−0.073	−0.061
(in %-points)	(0.027)	(0.029)	(0.027)	(0.029)	(0.022)	(0.025)
Elasticity of hours worked	−0.154	−0.167	−0.154	−0.167	−0.174	−0.161
(in %)	(0.048)	(0.050)	(0.048)	(0.050)	(0.049)	(0.049)
Own wage						
Change in participation	0.428	0.471	0.428	0.471	0.313	0.308
(in %-points)	(0.044)	(0.054)	(0.044)	(0.054)	(0.036)	(0.041)
Elasticity of hours worked	1.09	1.11	1.09	1.11	1.20	1.19
(in %)	(0.10)	(0.13)	(0.10)	(0.13)	(0.14)	(0.17)
Policy effects						
Change in participation	−2.56	−2.96	−2.56	−2.96	−3.26	−3.13
(in %-points)	(0.38)	(0.48)	(0.38)	(0.48)	(0.55)	(0.49)
Change in hours worked	4.19	4.20	4.19	4.20	10.15	10.50
(in %)	(0.67)	(0.79)	(0.67)	(0.79)	(1.22)	(1.39)
Likelihood	−8183.80	−8137.02	−8183.80	−8137.02	−9466.96	−9436.71
	(94.0)	(92.7)	(94.0)	(92.7)	(68.1)	(70.2)
AIC	9.15	9.12	9.15	9.12	10.58	10.56
	(0.10)	(0.10)	(0.10)	(0.10)	(0.076)	(0.078)

**Note*: 100 data sets are generated using fifth order model estimates (cf. Table 5). Means and standard deviations (in parentheses) over these 100 are reported; measurement error is added to hours worked with either standard deviation equal to 2 or standard deviation equal to 4.

but still modest. The conclusion seems clear: the performance of the simulated ML estimator in case of no misspecification is remarkably good. We find no evidence of either a finite sample bias, or a bias due to small R.

The first column of Table 3 shows the bias due to using a second order specification of preferences rather than the fifth order specification. The differences with column 2 and with the true values given in Table 2 are rather small. They are similar to the differences obtained when the real data are used, see the fourth panel in Table 2. Thus although the fifth order model is the true model and the second order model is misspecified, the bias on the elasticities and the policy effects induced by this misspecification is very limited. This conclusion obviously might be specific to our empirical example, where the second order approximation works reasonably well.

We repeated this exercise for the first, third and fourth order model. Results are not presented to save space, but they all confirm the findings in Table 2. The third and fourth order models give similar results as the fifth and second order models. The first order model, however, appears to be too restrictive for the data at hand, and leads to grossly biased estimates of elasticities and policy effects. This should not come as a surprise, since linear utility is extremely restrictive.

The bottom panel of Table 3 shows the average log likelihood and AIC criterion values and their "sample" standard deviations over the 100 new data sets. On average, the likelihood of the fifth order model without measurement error in hours worked is substantially larger than that of the second order model. It appears that for 99 out of the 100 new data sets, a likelihood ratio test rejects the second order model in favour of the fifth order model. Similarly, although the difference is always small, the AIC value of the second order model is worse (i.e. larger) than the AIC value of the fifth order model in (the same) 99 data sets. This confirms the conclusion from Table 1 that from a statistical point of view, the fifth order model outperforms all other models. From an economic point of view, however, the conclusion is different: Tables 2 and 3 show that the second order model already reproduces the relevant elasticities and policy effects.

4.1. Measurement errors in hours worked

The traditional Hausman model has two types of errors in the labour supply equation: random preferences, and optimisation or measurement (of hours worked) error. In our labour supply model, random preferences are incorporated, and the GEV I errors could be seen as alternative specific utility evaluation errors, i.e. a form of optimisation error. They cannot be seen as measurement error on (desired) hours worked, however. To investigate whether neglecting measurement error on hours worked could bias the results we have generated new data sets including such measurement error, and re-estimated the model with these new data.

It is not clear what would be a reasonable size of the measurement error. Estimates of the Hausman model do not lead to common findings on this point. Since different questions are used in different surveys (annual hours versus weekly hours, for example), there is not much reason to expect that measurement errors always have the same order of magnitude. In the simulations below, we used measurement errors with mean zero and standard deviations 2 and 4 h per week.

The new data sets including measurement errors on hours worked are constructed from the data sets used for the other Monte Carlo simulations. First, hours without measurement error are generated from the set $\{0, 1/6, 2/6, \ldots, 60\}$, as discussed above. Then a measurement error drawn from $N(0,4)$ of $N(0,16)$ is added, then hours are again rounded to a multiple of $1/6$. The measurement error also affects the gross wage rates, which in the real as well as the simulated data are computed as the ratio of gross earnings and hours worked. As before, we only considered the second and fifth order models. The results are presented in the right hand columns of Table 3. Again, the elasticities and policy effects can be compared with the true values used to generate the data, given in the top panel of Table 2. They can also be compared to the figures in the first two columns of Table 3, based upon the same 100 new data sets without measurement errors.

We find that the wage elasticities of desired working hours are not affected very much by the measurement errors. The results for the effects on participation, however, lead to some concern. If the measurement error is substantial (with standard deviation 4 hours per week), the estimated effects on participation of changing the women's or their husbands' wages, are strongly biased towards zero. A closer look at the parameter estimates suggests that this problem is probably related to identifying fixed costs. Fixed costs estimates for the data with measurement error, lead to a large number of people with negative predicted fixed costs, while fixed costs were almost always positive according to the earlier estimates. The reason is that the measurement error reduces the gap in the hours distribution at small part-time jobs.

The policy effect is another non-linear and complicated function of the parameters. It is related to the elasticities in the sense that both are driven by the sensitivity of labour supply to financial incentives. Still, the estimates of the policy effects are not in line with the results for the elasticities. The effect of the tax reform on participation is overestimated, though not dramatically. The effect of the tax reform on total hours worked is grossly overestimated by more than 100%. Thus measurement error in hours worked can have a detrimental effect on the estimates of the policy relevant parameters.

5. Conclusions

In this paper we have shown how flexible, non-parametric features can be built into a fully structural econometric model that is useful for policy analysis. We have combined a framework with a rich economic structure (utility maximisation under a complex budget constraint) with a non-parametric specification of the key element in this framework (the direct utility function). We have taken a static neo-classical labour supply model, one of the most popular structural models in (cross-section) econometrics over the past two decades. We have replaced the direct utility function by a flexible polynomial expansion, able to accurately approximate any utility function in a given compact set of relevant hours income combinations. Using the direct utility function only is made possible by treating the labour supply decision as a discrete choice problem, approximating the budget frontier by a finite set. The model also deals with many other problems in the structural labour supply literature, like non-participation, fixed costs of working, unobserved wage rates, measurement error in wage rates, and model coherency.

Using data for women with partner in the Netherlands, we use the framework to investigate female labour supply elasticities and the consequences of a recently proposed reform of income taxes, which implies disincentives for small part-time jobs and incentives to work full-time. We find that the fifth order model statistically outperforms the lower order models, but fifth order and lower order models give almost the same outcomes for the elasticities and the policy effects.

We then proceed with some Monte Carlo simulations. We find that, if the model is correctly specified, the performance of our estimator is excellent: there is no finite sample bias on the elasticities and policy effects, and the asymptotic standard errors are

Table 4

Variable	Definition	Mean	S.D.
w_m	Gross hourly wage rate, husband	31.44	12.44
h_m	Hours worked per week, husband	41.71	9.44
w	Gross hourly wage rate, wife	13.77	13.97
h	Hours worked per week, wife	14.54	15.14
wh	Wife's earnings	334.59	396.6
$w_m h_m$	Husband's earnings	1296.83	596.7
oth	Other income (excl. child benefits)	27.29	118.3
	Dummies wife's education level[a]		
dedu2	Lower vocational training	0.296	0.444
dedu3	Intermediate vocational training	0.425	0.494
dedu4	High school	0.150	0.357
dedu5	Higher vocational training	0.021	0.142
dedu6	University degree	0.026	0.158
Age	Age of wife (in years)	38.22	8.87
nkid18	Number of children age 0–18	1.19	1.17
dkid05	1 if child younger than 6 is present	0.278	0.448
chbf	Child benefits (Dfl per week)	50.80	47.45

[a]Reference category: primary education only.

very good approximations of the true standard errors. This finding seems in contrast with many earlier Monte Carlo Studies (such as Blomquist, 1996), though it should be realised that our sample size is relatively large (1794 observations, versus 602 in Blomquist, 1996). The conclusion from the final set of Monte Carlo is less positive: measurement error on hours of work can bias the estimates of policy effects.

The model considered here is obviously just one example. Numerous directions of extension exist. Some of these can straightforwardly be incorporated in the current framework. The economic model can be extended to incorporate joint decisions of husband and wife, gross wage rates varying with hours worked, or costs of childcare. For example, the assumption that the husband's and the wife's personal incomes are perfect substitutes in the wife's utility function is quite restrictive and has been empirically rejected (see Kooreman, 2000). Econometric extensions include allowing for more flexible distributions of error terms in the wage equation or in random preferences, or explicitly allowing for measurement error. The current results suggest that even with a limited number of observations such extensions can be useful.

Acknowledgements

We are grateful to Amos Golan, Francois Laisney and two anonymous referees for useful comments. Statistics Netherlands is acknowledged for providing the data.

Table 5

	Estimates	t-value
Utility function		
y^5	0.0354	0.13
$y^4 * (h/10)$	0.0062	0.04
$y^3 * (h/10)^2$	−0.104	−0.64
$y^2 * (h/10)^3$	−0.0056	−1.05
$y * (h/10)^4$	−0.0207	−0.95
$(h/10)^5$	0.0354	6.04
y^4	−0.0819	−0.05
$y^3 * (h/10)$	0.252	0.23
$y^2 * (h/10)^2$	1.29	1.54
$y * (h/10)^3$	0.394	1.37
$(h/10)^4$	−0.504	−6.31
y^3	−1.21	−0.20
$y^2 * (h/10)$	−3.24	−0.96
$y * (h/10)^2$	−4.06	−2.34
$(h/10)^3$	2.53	5.66
y^2	7.11	0.68
$(h/10)^2$	−4.48	−3.22
$y * (h/10)$	4.49	1.90
y	2.45	0.28
$(h/10)$	3.33	1.29
$(h/10)*$nkid18	−0.518	−11.9
$(h/10)*$dkid05	−0.538	−5.42
$(h/10) * (\text{age}/10)$	−0.227	−0.57
$(h/10) * (\text{age}/10)^2$	−0.0391	−0.77
Fixed costs		
Constant	0.982	5.43
nkid18	−0.0334	−4.28
dkid05	−0.0062	−0.38
age/10	−0.165	−2.33
$(\text{age}/10)^2$	0.0179	2.08
Wage equation		
Constant	1.90	7.87
Age/10	0.464	3.74
$(\text{age}/10)^2$	−0.0570	−3.62
dedu2	0.0550	1.37
dedu3	0.242	6.12
dedu4	0.462	9.98
dedu5	0.598	8.78
dedu6	0.421	5.41
σ_{rp}	0.000	0.00
σ_u	0.197	9.98
σ_e	0.404	34.6

Appendix A.

Variable definitions and summary statistics, and estimates of the fifth order model is shown in Tables 4 and 5.

References

Aaberge, R., Colombino, U., Strøm, S., 1999. Labor supply in Italy. An empirical analysis of joint household decisions, with taxes and quantity constraints. Journal of Applied Econometrics 14, 403–422.

Blomquist, S., 1983. The effect of income taxation on the labour supply of married men in Sweden. Journal of Public Economics 22, 169–197.

Blomquist, S., 1995. Restrictions in labor supply estimation: is the MaCurdy critique correct? Economics Letters 47, 229–235.

Blomquist, S., 1996. Estimation methods for male labour supply functions: how to take account of non-linear taxes. Journal of Econometrics 70, 383–405.

Blomquist, S., Newey, W., 1997. Nonparametric estimation of labor supply functions generated by piece wise linear budget constraints. Working Paper No. 1997-24, Uppsala University.

Blundell, R., MaCurdy, T., 1999. Labor supply: a review of alternative approaches. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3A. North-Holland, Amsterdam, pp. 1559–1695.

Blundell, R., Walker, I., 1986. A life cycle consistent empirical model of labour supply using cross-section data. Review of Economic Studies 53, 539–558.

Burtless, G., Hausman, J., 1978. The effect of taxes on labor supply. Journal of Political Economy 86, 1103–1130.

Callan, T., van Soest, A., 1996. Family labour supply and taxes in Ireland. Mimeo, Tilburg University.

Dickens, W., Lundberg, S., 1993. Hours restrictions and labor supply. International Economic Review 34, 169–192.

Euwals, R., van Soest, A., 1999. Desired and actual labour supply of unmarried men and women in the Netherlands. Labour Economics 6, 95–118.

Gabler, S., Laisney, F., Lechner, M., 1993. Seminonparametric estimation of binary-choice models with an application to labor-force participation. Journal of Business & Economic Statistics 11, 61–80.

Gallant, R., Nychka, D., 1987. Semi-nonparametric maximum likelihood estimation. Econometrica 55, 363–390.

Golan, A., 1988. A discrete stochastic model of economic production and a model of fluctuations in production—theory and empirical evidence. Ph.D. Thesis, University of California, Berkeley.

Golan, A., Judge, G., Miller, D., 1996. Maximum Entropy Econometrics. Wiley, Chichester.

Gong, X., van Soest, A., 2002. Female labour supply and family structure in Mexico City. Mimeo, Tilburg University, Journal of Human Resources 37, forthcoming.

Hajivassiliou, V., Ruud, P., 1994. Classical estimation methods for LDV models using simulation. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, Vol. IV. North-Holland, New York, pp. 2384–2443.

Hausman, J., 1981. Labor Supply, in How Taxes Affect Economic Behavior. Brookings Institution, Washington DC.

Hausman, J., 1985. The Econometrics of nonlinear budget sets. Econometrica 53, 1255–1283.

Heckman, J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.

Heckman, J., Singer, B., 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52, 271–320.

Keane, M., Moffitt, R., 1998. A structural model of multiple welfare program participation and labor supply. International Economic Review 39, 553–589.

Killingsworth, M., Heckman, J., 1986. Female labor supply: a survey. In: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics, Vol. 1, pp.103–204. North-Holland, Amsterdam.

Kooreman, P., 2000. The labeling effect of a child benefit system. American Economic Review 90, 571–583.

MaCurdy, T., Green, D., Paarsch, H., 1990. Assessing empirical approaches for analyzing taxes and labor supply. Journal of Human Resources 25, 415–490.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete choice. Journal of Applied Econometrics 15, 447–470.

Ministry of Finance, 1997. Belastingen in de 21-ste eeuw: een verkenning. SDU Publishers, The Hague.

Ministry of Finance, 1999. Wet inkomstenbelasting 2001. SDU Publishers, The Hague.

Moffitt, R., 1984. Estimation of a joint wage-hours labor supply model. Journal of Labor Economics 2, 550–556.

Moffitt, R., 1986. The econometrics of piecewise linear budget constraints: a survey and exposition of the maximum likelihood method. Journal of Business and Economic Statistics 4, 317–327.

Moffitt, R., 1990. The econometrics of kinked budget constraints. Journal of Economic Perspectives 4, 119–139.

Nakamura, A., Nakamura, M., 1990. The econometrics of female labor supply and children. Econometric Reviews 11, 1–71.

Pollak, R., Wales, T., 1979. Welfare comparisons and equivalence scales. American Economic Review, Papers and Proceedings 69, 216–221.

Soofi, E., 1992. A generalizable formulation of conditional logit with diagnostics. Journal of the American Statistical Association 87, 812–816.

Theeuwes, J., 1988. Arbeid en Belastingen, in Belastingheffing en Belastinghervorming, Koninklijke Vereniging voor Staathuishoudkunde, Stenfert Kroese, Leiden, 111–143.

Tummers, M., Woittiez, I., 1991. A simultaneous wage and labour supply model with hours restrictions. Journal of Human Resources 26, 393–423.

Van Soest, A., 1995. Discrete choice models of family labor supply. Journal of Human Resources 30, 63–88.

Van Soest, A., Woittiez, I., Kapteyn, A., 1990. Labour supply, income taxes and hours restrictions in the Netherlands. Journal of Human Resources 25, 517–558.

Van Soest, A., Kooreman, P., Kapteyn, A., 1993. Coherency and regularity of demand systems with equality and inequality constraints. Journal of Econometrics 57, 161–188.

Van Soest, A., Das, M., 2001. Family labour supply and proposed tax reforms in the Netherlands. De Economist 149, 191–218.

Vlasblom, J.D., 1998. Differences in labour supply and income of women in the Netherlands and the Federal Republic of Germany. Ph.D. Thesis, University of Utrecht.