Different Modeling Strategies for Discrete Choice Models of Labor Supply: Estimates for Switzerland

Reto Nyffeler
University of Bern

Incomplete Draft, Please Do not Quote
April 2005

Abstract

In recent applications of discrete choice models of labor supply considerable attention has been devoted to strategies to increase the flexibility of models for a better fit to the data. However the functional form of preferences in these models remained restrictive. The question is therefore if we can gain something by allowing for more flexible preferences. This paper compares four different modeling strategies: (1) A structural model with fixed costs of work and random heterogeneity (2) A model with a nonparametric specification of the direct utility function. (3) A model which allows parameters to be fully alternative specific and which remains agnostic about the interpretation of its parameters (4) A model that is equivalent to the one in (3) but allows for price and income dependent preferences. Results show a clear rejection of the restrictions imposed in the structural model. Moreover estimation of the model with price and income dependent preferences lead to a clear rejection of the standard unitary approach. However model (3) and especially (4) cause a significant portion of the sample to not respect the needed regularity condition (utility increasing in income). Based on the estimates of models (1) to (3) the introduction of a flat rate tax is simulated and the estimated elasticities of the models are compared.

Correspondence:
Reto Nyffeler
Department of Economics
University of Bern
Gesellschaftsstr. 49
3012 Bern
Switzerland
reto.nyffeler@vwi.unibe.ch

1 Introduction

Over the last decade the discrete choice approach of labor supply analytics gained more and more popularity in assessing the impact of public policies on work incentives. The approach can easily handle non-linear and possibly non-convex budget sets caused by taxation. In addition it avoids the MaCurdy, Green, and Paarsch (1990) critique that coherency of the model implicitly limits the range of elasticities that can be obtained.

Recent studies of discrete labor supply focused very much on improving the models capability in explaining the peaks in the hours distribution. This was done by gaining flexibility through the introduction of random parameters, state specific constants or fixed cost of work (see for example (Van Soest 1995)). However the functional form of preferences in these models remained restrictive. Since discrete choice analytics does not need explicit expressions for both the direct utility function and the labor supply function (or the indirect utility or expenditure function), very general functional forms of preferences are principally possible. More fundamentally there is strong evidence against the standard or unitary approach within most policy analysis is done. The question is therefore if we can gain something by allowing for more general preferences.

In the present paper we test a structural model with fixed cost of work and random heterogeneity against models that allow for more general preferences. These more general models are taken from two recent contributions, one by Van Soest, Das, and Gong (2000) and the other by Bargain (2004). In the framework of Van Soest et al. the direct utility function is approximated by a nonparametric series approximation in hours and income. In this way they introduce a structural nonparametric labor supply model which can be used for all sorts of policy analysis. The parameters of the model remain interpretable. Bargain suggests two generalizations of the structural model that are more radical than the one from Van Soest et al.. An interpretation of any of the parameters is not possible. In the first suggestion preference parameters are allowed to be alternative specific, that is utility can depend on disposable income in a fully flexible way over working hours. The second generalization allows the utility of each alternative to depend on disposable income as well as on wage rates and non-labor income. In some sense then preferences are price and income dependent. Since this model does not verify Slutsky conditions or pooling it is not of the unitary type. Hence the models allows a test of the unitary approach. The model can be motivated in different ways including the collective approach, the life cycle framework and demand side aspects. However it does not allow to discriminate between the different approaches. Despite their generality all the flexible models discussed in this paper maintain a utility maximizing interpretation.

Finally based on the estimates of the models the introduction of a flat tax rate is simulated and the estimated elasticities of the different models are compared.

The paper is structured as follows: Section 2 presents an overview of the logit models used to represent the utility maximizing behavior of the decision makers. In section 3 the specification of the different models is introduced and discussed. In section 4 the performance of different models with respect to explanatory power and consistency with economic theory is analyzed. Section 5 presents the structure of the reform and analysis the predicted labor supply responses of the different models. Section 6 concludes.

2 Logit Models for Multiple Choices

In discrete choice labor supply modeling the labor supply decision is described as the utility maximizing choice between discrete hours alternatives. A prominent way to model utility maximizing discrete choices are random utility models (RUMs). They are the basis of all the choice models used in this paper and can be derived as following.[1] A decision maker i faces a choice among J alternatives. Each alternative provides a certain level of utility. From alternative J the decision maker obtains utility U_{ij}, $j = 1, ..., J$. Alternative j is chosen if $U_{ij} > U_{ik}$ for all $k \neq j$. The decision maker's utility can be decomposed as

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \tag{1}$$

where V_{ij} is a function which relates observed factors to the decision maker's utility. These factors are attributes of the alternatives, X_{ij} $\forall j$, and some attributes of the decision maker, S_i. V_{ij} depends on unknown parameters β_j which have to be estimated. The function is denoted $V_{ij} = V(X_{ij}; S_i, \beta_j)$ $\forall j$ and is called representative utility. Factors that are not included in V_{ij} but affect utility are captured by ε_{ij}. This part of the utility is unknown and assumed to be random. It can be seen as the error made in evaluating alternative j. Since ε_{ij} is simply the difference between U_{ij} and V_{ij} this decomposition is completely general.

The logit model is obtained by assuming that each ε_{ij} is independently, identically distributed extreme value. The density and cumulative distribution of ε_{ij} are respectively

$$f(\varepsilon_{ij}) = e^{-\varepsilon_{ij}} e^{-e^{\varepsilon_{ij}}}$$

and

$$F(\varepsilon_{ij}) = e^{-e^{-\varepsilon_{ij}}}.$$

Mc Fadden (1974) has proved that under this assumption the probability that decision maker i chooses alternative k is

$$\begin{aligned} P_{ij} &= Prob(V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij} \ \forall k \neq j) \\ &= \frac{e^{V_{ij}}}{\sum_k e^{V_{ik}}} \end{aligned} \tag{2}$$

[1]See Train (2003) for an excellent overview.

Representative utility can either be specified to be linear or nonlinear in parameters. If parameters enter representative utility nonlinearly estimation is more difficult because the log-likelihood function may not be globally concave.

Since the logit probabilities take a closed form, the traditional maximum-likelihood procedures can be applied. The log-likelihood function is given by

$$LnL(\beta_j) = \sum_{i=1}^{I} \sum_{i} d_{ij} \; lnP_{ij}, \tag{3}$$

where $d_{ij} = 1$ if person i chose j and zero otherwise. The simplicity of the logit model is a strong advantage. But logit models have some clear limitations. They can only represent systematic taste variation but not random taste variation and they imply proportional substitution across alternatives, that is logit models exhibit the IIA property. One model that obviates these disadvantages is the mixed logit model (Brownstone and Train 1998, McFadden and Train 2000).

The mixed logit choice probability can be derived in several ways form utility maximizing behavior. The following derivation is based on the random coefficient interpretation (Revelt and Train 1998).[2] The utility of person i from alternativ j is given by

$$U_{ij} = V_{ij} + \varepsilon_{ij} = V(X_{ij}; S_i, \beta_i) + \varepsilon_{ij}, \tag{4}$$

where X_{ij}, S_i and ε_{ij} are defined as before and β_i is a vector of coefficients for person i.[3] If utility is linear in β_i and we abstract from S_i utility can be written as $U_{ij} = \beta_i' x_{ij} + \varepsilon_{ij}$. This specification is the same as for logit, except that now the coefficients β_i vary randomly over the decision maker rather than being fixed. The coefficient vector for each decision maker can be expressed as the sum of the mean, b, and individual deviation, η_i. Utility is then $U_{ij} = b' x_{ij} + \eta_i' x_{ij} + \varepsilon_{ij}$. The unobserved portion of utility is $\eta_i' x_{ij} + \varepsilon_{ij}$. This term is correlated over alternatives due to the common η_i. Because of this correlation, mixed logit does not exhibit the independence from irrelevant alternatives property.[4] If we knew the decision maker's taste, that is, if we knew the value of β_i the conditional choice probability would be standard logit since ε_{ij}'s are iid extreme value, that is

$$L_{ij}(\beta_i) = \frac{e^{V_{ij}(\beta_i)}}{\sum_{k=1}^{J} V_{ik}(\beta_i)}$$

[2]An other popular interpretation is based on error components. But since here the stress is more on individual taste variation and less on substitution patterns the random coefficient interpretation seems more natural.

[3]For notational simplicity we use here β_i instead of β_{ij}. However it is no problem to generalize mixed logit to allow for alternative specific random coefficients. In order to avoid the IAA property either the variance of these random coefficients have to be the same for all alternatives or the random coefficients are allowed to be correlated over alternatives.

[4]Mixed logit allow for very general patterns of correlation and hence very general patterns of substitution. McFadden and Train (2000) have shown that any random utility model can be approximated by mixed logit.

Since β_i is not given (we can estimate b but can not observe η_i for each decision maker), the (unconditional) choice probability is this logit formula integrated over all values of β_i

$$P_{ij} = \int \frac{e^{V_{ij}(\beta)}}{\sum_{k=1}^{J} V_{ik}(\beta)} f(\beta) d\beta, \tag{5}$$

Models of this form are called mixed logit because the choice probability is a mixture of logits with $f(\beta)$ as the mixing distribution.[5] The mixing distribution may be discrete or continuous. In the discrete case the mixed logit becomes the so called latent class model.[6] As most applications of mixed logit we assume the density of β to be continuous and more specific to be normal with mean b and covariance W. In this case the choice probability is given by

$$P_{ij} = \int \frac{e^{V_{ij}(\beta)}}{\sum_k e^{V_{ik}(\beta)}} \phi(\beta|b, W) d\beta, \tag{6}$$

where $\phi(\beta|b, W)$ is the normal density with mean b and covariance W. The parameters to be estimated are those of the mixing distribution $f(\beta)$, b and W.

Since there is no closed form expression for the choice probabilities in mixed logit we approximate the probabilities by simulation and maximize the simulated log-likelihood function. In particular for given b and W a value of β is drawn from $f(\beta|b, W)$. This value is labeled β^r with the superscript r=1 referring to the first draw. Using this draw the standard logit formula $L_{ij}(\beta^r)$ is calculated. This process is repeated for many draws and the results are averaged. This average is the simulated probability:

$$\check{P}_{ij} = \frac{1}{R} \sum_{r=1}^{R} L_{ij}(\beta^r) \tag{7}$$

where R is the number of draws. $\check{P}_{ij}$ is an unbiased estimator of P_{ij}. Its variance decreases as R increases. The simulated probabilities are inserted into the log-likelihood function to give a simulated log-likelihood:

$$SSL = \sum_{i=1}^{I} \sum_{j=1}^{J} d_{ij} ln \check{P}_{ij}, \tag{8}$$

where d is defined as above. The maximum simulated likelihood estimator is the value of b and W that maximizes the simulated log likelihood.

[5] The standard logit model is a special case where $f(\beta_j)$ is degenerate at fixed parameters b: $f(\beta) = 1$ for $\beta = b$ and 0 for $\beta \neq b$.

[6] In the latent class model $\beta = b_m$ with probability s_M. The choice probability for this model is given by

$$P_{ij} = \sum_{m=1}^{M} s_m \frac{e^{V_{ij}(b_m)}}{\sum_k e^{V_{ik}(b_m)}}$$

Here it assumed that the population consists of M segments each with its own choice behavior or preferences. The share of segment m in the population, s_m, is estimated along with the b_m's for each segment (see Bargain (2004) for a recent application to labor supply estimation.

3 Specification of the Models

3.1 Structural Models of Labor Supply

We describe a static neo-classical structural lobor supply model for single decision makers. Following Keane and Moffitt (1998) and Blundell, Duncan, McCrae, and Meghir (2000) we assume that the answer to the desired hours question is based upon maximizing

$$U_{ij} = V(Y_{ij}, H_j; S_i, \beta) + \varepsilon_{ij}, \tag{9}$$

where Y_{ij} is net household income, H_j are female hours of work and S_i are household characteristics. The net household income Y_{ij} is given by

$$Y_{ij} = w_i H_j + Y_{im} + Y_{inl} - T(w_i H_j, Y_{im}, Y_{inl}; S_i)$$

where w_i is the female's wage rate, Y_{im} the husband's labor income, Y_{inl} is the household's non labor income and $T(w_i H_j, Y_{im}, Y_{inl}; S_i)$ are the tax payments. As in Blundell, Duncan, McCrae, and Meghir (2000) the utility function is specified to be quadratic and is given by

$$U_{ij} = \beta^{YY} Y_{ij}^2 + \beta^{HH} H_j^2 + \beta^{YH} Y_{ij} H_j + \beta^{Y} Y_{ij} + \beta^{H} H_j \quad \text{for } j = 1, ...J \tag{10}$$

observed heterogeneity is introduced by assuming that

$$\beta^H = \beta_{h0} + \beta_h' S_i. \tag{11}$$

In principle there is no theoretical reason to only allow β^H to vary with X. However the identification of the effects of X via different β's is often difficult. In addition β^H is an attractive choice for interpreting the results. It implies that the marginal utility of work varies linearly with X. The sign of the β_h coefficients directly determines if the variables in X have a positive (positive sign) or a negative (negative sign) effect on the marginal utility of work. This basic model will hereafter be referred to as model S1. It has two major shortcomings. First, it does not fit the data, in the sense that it underpredicts nonparticipation and overpredicts part-time jobs involving a few hours a week. Second, it does not allow for unobserved individual heterogeneity. Several methods have been used to overcome the first shortcoming. Van Soest (1995) introduced some hours specific constants on an ad hoc basis in the utility function. These may reflect costs of finding a part-time job. An alternative with a more attractive economic interpretation is the incorporation of fixed costs of work (Callan and Soest 1996).[7] Fixed costs are the costs an individual has to pay to get to work. By subtracting them from income for the strictly positive working hours they can be introduced into the model in a natural way. For countries with very high costs of childcare like Switzerland fixed costs of work are mainly made up

[7] Another alternative would be the approach of Dickens and Lundberg (1993), who incorporate demand-side restrictions on hours worked explicitly, but this model requires strong assumptions for identification.

by childcare costs. In principal it would therefore be preferable to proceed as Blundell, Duncan, McCrae, and Meghir (2000) and use sample information on hourly prices of childcare to account for chilcare expenditures. However data on childcare costs for Switzerland are of a too poor quality. We therefore have to define a fixed costs equation in terms of a set of observable variables. Fixed costs are assumed to be not stochastic and are specified as

$$F_i = \delta' Z_i, \tag{12}$$

where Z_i is a subset of S_i.[8] For all states $j > 0$ utility expression 9 becomes

$$U_{ij} = V(Y_{ij} - F_i, H_j; S_i, \beta) + \varepsilon_{ij}, \tag{13}$$

If utility increases with income, fixed costs decrease the utility of working, thereby increasing the probability of nonparticipation. This model will be referred to as model S2. The second shortcoming can be removed by adding an error term to one of the parameters of the utility function. We follow this strategy and assume that unobserved heterogeneity enters through the parameter β^Y.[9]

$$\beta_i^Y = \beta_{y0} + \beta_y S_i + v_{iy}, \tag{14}$$

where $v_{iy} \sim N(0, \sigma^2_{v_{iy}})$. The model with fixed costs of work and random preferences will be referred to as model S3.

In contrast to the continuous labor supply model imposing Slutsky conditions (quasi-concavity of the direct utility function) is not necessary in discrete choice analysis (Van Soest, Kapteyn, and Kooreman 1993). Quasi-concavity of the utility function can be checked ex post. In this way the MaCurdy critique that elasticities are largely determined a priori (through the quasi-concavity restriction) can be avoided. If the utility function turns out not be quasi-concave the economic interpretation of the model is not affected since the interpretation depends not on concavity. The only restriction required for economic interpretation is that utility has to be increasing with income. This restriction we need since we assume that everyone always chooses a point on the frontier of the budget set rather than in the interior. However we will not impose this condition a priori before estimation but check it ex post (Van Soest, Das, and Gong 2000). This considerations are also valid for the more flexible models which follow in the next sections.

3.2 More Flexible Models of Labor Supply

The improvements of model S1 through the introduction of fixed costs and random coefficients are strategies to increase flexibility of structural models for a

[8] We also experimented with stochastic fixed costs. However this did not help to improve the fit of the model.

[9] This choice is driven by the fact that the structural models must be comparable with the more flexible models where only the income terms remain in the utility function. However the model was also estimated with unobserved heterogeneity entering through β^H. The fit of the model did not improve at all.

better fit to the data. However the specification of preferences in these models is still restrictive. Discrete choice labor supply would allow for more flexible specifications since in contrast to continuous models discrete choice analytics does not need explicit expressions for both the direct utility function and the labor supply function (or the indirect utility or expenditure function).[10] The question is therefore if it is reasonable and possible to use more flexible specifications of preferences. We consider two recent contributions put forth by Bargain (2004) and Van Soest, Das, and Gong (2000) respectively which propose more flexible models of labor supply.

3.2.1 A Structural Labor Supply Model with Nonparametric Preferences

Van Soest, Das, and Gong (2000) introduce a structural nonparametric labor supply model (hereafter SNP model). Basically they replace the direct utility function of the structural model S3 with a flexible polynomial expansion. In this way they maintain the economic structure of model S3 (utility maximization under a complex budget set) and combine it with a nonparametric specification of the utility function. Since the general structure of this flexible model remains unchanged it can be represented by expression 13:

$$U_{ij} = V(Y_{ij} - F_i, H_j; S_i, \beta) + \varepsilon_{ij},$$

The direct utility function is specified as higher order polynomial in its arguments H and Y:

$$U_{ij} = \sum_{p=0}^{K} \sum_{q=0}^{K-p} \beta^{Y^p H^q} H^p (Y - F)^q \tag{15}$$

K is the order of the polynomial and determines the flexibility of the utility function. Since for K equal to two we get the model discussed in the previous section K has to be larger than two. If K is allowed to be arbitrarily large, U_{ij} is able to approximate any utility function in a given compact set of relevant hours income combinations. However for finite sample size the order of the polynomial that can be used is limited.[11] In the empirical section we will consider the case of K=5 the largest value of K Van Soest et al. used in their work. As in the previous section observed and unobserved heterogeneity is assumed to enter through the parameters β^Y and β^H

$$\begin{aligned} \beta_i^Y &= \beta_{y0} + v_{iy} \\ \beta^H &= \beta_{h0} + \beta_h' S_i, \end{aligned}$$

where $v_{iy} \sim N(0, \sigma^2_{v_{iy}})$. Fixed costs are specified and introduced into the model as before. An important comment can be made concerning the identification of

[10] See for example Creedy and Duncan (2002).

[11] Asymptotics requires that K tends to infinity much slower than the number of observations.

fixed costs. The identification of fixed costs can be intuitively explained by the lack of observations with low working hours. In the case of a fully nonparametric utility function it could be that the utility function itself could pick up the gap in the distribution at low hours, by assigning lower utility to such hours values. Fixed costs would then be nonparametrically unidentified. However since X enters the utility function and fixed costs in a restrictive way this should not be a matter of concern here.

3.2.2 Unconstrained and Non-standard Models of Labor Supply

Bargain (2004) suggests two generalizations of model S3 which relax the restrictions on household preferences imposed in this model step by step. His generalizations are more radical than the one from Van Soest, Das, and Gong (2000). However despite their generality both models maintain a utility maximizing interpretation.

Unconstrained Model In this model preference parameters are alternative specific, that is utility can depend on consumption in a fully flexible way across working hours. A direct interpretation of any of the parameters is no more possible. The model which nest the structural models from section 3.1 is given by

$$U_{ij} = V(Y_{ij}; S_i, \beta_j) + \varepsilon_{ij}, \tag{16}$$

where Y_i is given as before.[12] Using the quadratic form the utility function for this model has the form

$$U_{ij} = \beta_j^{YY} Y_{ij}^2 + \beta_j^Y Y_{ij} + \delta_j \quad \text{for } j = 1, ...J \tag{17}$$

observed and unobserved heterogeneity is written as

$$\begin{aligned} \beta_j^Y &= \beta_{y0j} + \beta'_{yj} S_i + v \\ \delta_j &= \delta_{0j} + \delta'_j S_i + \sum_{k=1}^{L} \sum_{l=k}^{L} \delta_j^{kl} s_i^k s_i^l \end{aligned} \tag{18}$$

Only $J-1$ sets of parameters δ_j can be identified. For the first alternative the δ coefficients are therefore set to zero. Since disposable income is alternative specific all J β coefficients can be estimated.

Non-standard Model So far wage rates and non-labor income influence labor choices only through disposable income. This is consistent with the standard or unitary approach and implies income pooling and common preferences within a household. Bargain's second generalization allows each alternative to depend on disposable income as well as on wage rates and non-labor income. In some

[12]See the appendix of Bargain (2004) for the identification fo the constraints imposed by model S3 on the unconstrained model

sense then preferences are price- and income-dependent. Such a non-standard model could have the following form:

$$U_{ij} = V(Y_{ij}, w_i, Y_{im}, Y_{inl}; S_i, \beta_j) + \varepsilon_{ij}, \tag{19}$$

where the variable definitions remain the same as above. Utility of alternative j is now not only dependent on disposable income of the household but also on female's wage rate, the labor income of the husband and non-labor income. This model can be rationalized in different ways. It can be related to the collective approach, it can be made consistent with the life cycle framework and finally the model could reflect constraints from the demand side.[13] However the model does not allow to discriminate between these different approaches. Keeping the quadratic form the utility function of the model could be specified as

$$\begin{aligned} U_{ij} \;=\; & \beta_j^{YY} Y_{ij}^2 + \beta_j^{Y} Y_{ij} + \beta_j^{ww} w_i^2 + \beta_j^{Y_m Y_m} Y_{im}^2 + \beta_j^{Y_{nl} Y_{nl}} Y_{inl}^2 \\ & + \beta_j^{wY_m} w_i Y_{im} + \beta_j^{wY_m} w_i Y_{inl} + \beta_j^{Y_m Y_{nl}} Y_{im} Y_{inl} + \beta_j^{w} w_i + \beta_j^{Y_m} Y_{im} \\ & + \beta_j^{Y_{nl}} Y_{inl} + \beta_j^{wY} w_i Y_{ij} + \beta_j^{Y_m Y} Y_{im} Y_{ij} + \beta_j^{Y_{nl} Y} Y_{inl} Y_{ij} + \delta_{ij} \\ & \text{for } j = 1, ...J \end{aligned} \tag{20}$$

with:

$$\begin{aligned} \beta_j^{Y} &= \beta_{y0j} + \beta_{yj}' S_i + v \\ \beta_j^{R} &= \beta_{r0j} + \beta_{rj}' S_i \quad \text{for} \quad r = w, Y_m, Y_{nl} \\ \delta_{ij} &= \delta_{0j} + \delta_j^{'} S_i + \sum_{k=1}^{L} \sum_{l=k}^{L} \delta_j^{kl} s_i^k s_i^l \end{aligned} \tag{21}$$

4 Data and empirical results

4.1 Data

The data used in this analysis are drawn from the Swiss Income and Expenditure Survey 1998 (SIES). Over 9000 households participated in this survey conducted by the Swiss Federal Office of Statistics. The survey is primarily used for the periodical revisions of the Swiss National Consumer Price Index. Besides the detailed expenditure data including tax and social security payments the survey also provides information about all sources of income as well as about labor supply of each household member.

For the empirical analysis we select married or de-facto couples, aged between 20 and 65, who are employed or voluntarily unemployed. Students, self employed, unemployed or retired people are excluded from the sample. Moreover people who work more than 60 hours a week, households with more than four children or with more than two decision makers are selected out. People with very high levels of non-labor income and individuals with wages below or

[13] Bargain (2004) provides a short description of the three approaches.

above the 1st and 99th percentiles of the wage distribution were also discarded. Since men's participation rate is very high (99.5%) and almost all men work full time the empirical analysis fully concentrates on female labor supply. Working hours of men are fixed at the observed value. With this selection the sample contains 2450 households. Table 1 displays descriptive statistics for the sample. Since gross wage rates for non-working individuals are not observed these wages are predicted using the standard Heckman two-step estimation procedure. For workers the actual wage rates are used.[14] Figure 1 displays the distributions of predicted and observed wage rates for part-time and full-time female workers respectively. The fit is more satisfying for full time workers than for part time workers. In both cases the predicted wage distribution is more concentrated around the mode. Figure 2 shows the distribution of female weekly working hours. A significant portion of females in couples does not participate in the labor market and the fraction of part-time working females is quite large. For the empirical analysis we assume that women have the following discrete choice set: H$\in \{0, 8, 16, 25, 33, 42\}$.

4.2 Empirical Results

Table 2 displays the estimation results for the structural models S1, S2 and S3. The interpretation of the parameters has to be made with caution. Directly interpretable are the interactions between hours worked and household characteristics. These coefficients determine how marginal utility changes with household characteristics. Age and the presence of children decrease marginal utility of work. In the case of children the effect is stronger for preschool children than for schoolaged children. High education increases the marginal utility of work. These results seem consistent with intuition and are in line with other studies (see for example Duncan and Harris (2002).

Fixed costs of work significantly decrease with the number of children. This counterintuitive result was also found by other studies (see for example Duncan and Harris (2002) and Van Soest, Das, and Gong (2000)). We estimated also a model in which the preschool coefficient of fixed costs could vary freely with alternatives. Results showed that the coefficient is only negative for small number of working hours and turned to be significantly positive for higher working hours.[15] These results may indicate that it is very attractive for women with small children to work for a small number of hours a week. On the whole the fixed cost coefficients are implausible high. Depending on the model average fixed costs represented by the constant term are more than 100% of the average earnings of working women. What exactly these coefficients measure is not clear.

[14]We are aware of the fact that this approach in principle does not lead to consistent estimators since it assumes that wage rates of nonworkers are predicted without errors . For consistent estimators it would be necessary to take the wage rate prediction errors explicitly into account for example by integrating out the disturbance term of the wage equation in the likelihood (Van Soest 1995).

[15]However this model heavily overpredicted non participation and was therefore dropped out

The inclusion of random preferences seems to considerably improve the precision of the preference parameters and leads to quite a large increase of the hours and fixed cost coefficients. Given the significantly estimated standard deviation of the distribution of the random coefficient there seems to exist considerable heterogeneity concerning income preferences.

The considerations just made about the interpretation of the parameters remain also valid for the SNP model. The estimation results for this model are displayed in table 3. Due to convergence problems this model is estimated without individual heterogeneity. Given the large number of parameters and the difficulty of interpretation the estimation results for the UC and GC model are omitted.[16] These models have been estimated with and without random parameter. However the introduction of individual heterogeneity did not improve the fit of the models at all. In the following comparison of the models we thus ignore individual heterogeneity.

Table 4 contains some information about the fit of the estimated models. In the upper part of the table the observed frequencies are compared with its average estimated value over all households. Not surprisingly the flexible models SNP, UC and GC perform best in this respect. However apart from the 8 and 33 hours alternative the probabilities predicted by the simplest model S1 are quite accurate as well. This may be due to the fact that in Switzerland female part-time work is widely spread and the pattern of working hours is not as rigid as in other countries. Another measure of fit displayed in table 4 is the pseudo-R2 or Likelihood Ratio Index of McFadden (1973). The measure is defined as $1 - LogL_e/LogL_0$, where $LogL_e$ is the log-likelihood function for the estimated model and $LogL_0$ is log-likelihood function when all parameters are set to zero. The definition of the measure implies that it is always between zero and one. According to this measure the general model clearly provides a better fit than the standard models and the flexible standard models dominate the simple structural model S1. Again this comes with no surprise.

Table 5 displays the log-likelihood values for the models S2, SNP, UC and GC. In addition it provides the LR statistics and the relevant critical values at the 1% significance level. Tests of model S2 against model SNP and UC result in a rejection of model S2 in both cases. This implies that from a statistical point of view the restrictions in the structural model S2 are too restrictive. Furthermore line three of table 5 shows that in a test of the UC model versus the GC model the standard model is rejected. This is nothing else than a test of the unitary model against a model with price-dependent preferences that does not verify Slutsky conditions or pooling (Pollak 1977). Thus this is strong evidence against the unitary approach.

As stated in section 3.1 the only coherency restriction we really need for the economic interpretation of the models is that utility is monotonically increasing in income. This restriction is satisfied for all observations and labor supply choices in the models S1, S2, S3 and SNP. In the models UC and GC however marginal utility of income is positive for only 85% and 1.7% of the labor supply

[16] Results available upon request from the author.

choices respectively. Thus the increased flexibility of model UC and GC has the advantage of capturing broader preference heterogeneity but has the disadvantage that a significant portion of the observations behave in contradiction to economic theory. In order to use the UC model in the simulation experiment we estimated the model again imposing the monotonicity restriction. Practically this was done by penalizing the log-likelihood for observations at which utility of a corresponding interior point of the budget set exceeds utility of the point on the edge. The resulting log-likelihood value is only slightly higher than the one from the unrestricted estimation. The log-likelihood value and the corresponding likelihood ratio are displayed on line four of table 5. Still the restrictions in model S2 are clearly rejected. For model GC we did not apply this procedure since the percentage of observations with positive marginal utility of income is too low. This model will not be used in the microsimulation.

5 Conclusions

This paper analyzes different modeling strategies for discrete choice labor supply models. The main result suggests that care should be taken when using very general functional forms of preferences in discrete choice labor supply analytics.

We compared four modeling strategies: a structural model with fixed costs of work and random heterogeneity, a model with a nonparametric specification of the direct utility function, a model which allows parameters to be fully alternative specific and a model that allows for price and income dependent preferences.

Some of the estimated parameters of the structural and the structural nonparametric model are directly interpretable. However as the implausible high coefficients of the fixed cost coefficients indicate the interpretation should be done very cautiously. What these coefficients exactly measure is unclear. Apart from fixed cost it could also be job search disutility, distaste of work or a mixture of all these. From this perspective the fact that none of the parameters of the unconstrained and non-standard model are interpretable is not a big disadvantage.

A series of likelihood ratio tests show that the restrictions made in the structural model are clearly rejected. In other words the structural nonparametric model as well as the unconstrained and the non-standard models are statistically superior to the structural model. Moreover estimation of the model with price and income dependent preferences lead to a clear rejection of the standard or unitary models.

However the unconstrained and non-standard model cause a significant part of the sample to not respect the only coherency restriction we really need for economic interpretation of the models and meaningful policy simulation: positive monotonicity in income. In the case of the non-standard model only 1.7% of the supply choices exhibit positive marginal utility of income. Improving results by a restricted estimation seems not to make sense here.

Overall the structural nonparametric model performs best. The model fits the data well and all supply choices exhibit positive marginal utility of income. In addition from an intuitive point of view it is not really obvious why the effect of income should differ over the supply choices as in the unconstrained and non-standard model. Furthermore it is not clear if these effects remain constant after the introduction of a tax reform.

References

Bargain, O. (2004): "On modeling household labor supply with taxation," DELTA Working Papers 2004-14, DELTA (Ecole normale suprieure).

Blundell, R., A. Duncan, J. McCrae, and C. Meghir (2000): "Evaluating In-Work Benefit Reform: The Working Families Tax Credit in the U.K," JCPR Working Papers 160, Northwestern University/University of Chicago Joint Center for Poverty Research.

Brownstone, D., and K. Train (1998): "Forecasting new product penetration with flexible substitution patterns," *Journal of Econometrics*, 89(1), 109–129.

Callan, T., and A. V. Soest (1996): "Family Labor Supply und Taxes in Irelnad," Discussion paper, Tilburg University.

Creedy, J., and A. Duncan (2002): "Behavioural Microsimulation with Labour Supply Responses," *Journal of Economic Surveys*, 16(1), 1–39.

Dickens, W. T., and S. J. Lundberg (1993): "Hours Restrictions and Labor Supply," *International Economic Review*, 34(1), 169–92.

Duncan, A., and M. N. Harris (2002): "Simulating the Behavioural Effects of Welfare Reforms among Sole Parents in Australia," *Economic Record*, 78(242), 264–76.

Keane, M., and R. Moffitt (1998): "A Structural Model of Multiple Welfare Program Participation and Labor Supply," *International Economic Review*, 39(3), 553–89.

MaCurdy, T., D. Green, and H. J. Paarsch (1990): "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply," *Journal of Human Resources*, 25(3), 415–90.

McFadden, D., and K. Train (2000): "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15(5), 447–70.

Pollak, R. A. (1977): "Price Dependent Preferences," *American Economic Review*, 67(2), 64–75.

Revelt, D., and K. Train (1998): "Mixed Logit With Repeated Choices: Households' Choices Of Appliance Efficiency Level," *The Review of Economics and Statistics*, 80(4), 647–657.

Train, K. E. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.

Van Soest, A. (1995): "Structural Models of Family Labor Supply: A Discrete Choice Approach," *Journal of Human Resources*, 30(1), 63–88.

Van Soest, A., M. Das, and X. Gong (2000): "A structural labour supply model with nonparametric preferences," Discussion paper, Tilburg University, Center for Economic Research.

Van Soest, A., A. Kapteyn, and P. Kooreman (1993): "Coherency and Regularity of Demand Systems with Equality and Inequality Constraints," *Journal of Econometrics*, 57(1-3), 161–88.

Table 1: Descriptive statistics

Variable	Women	Men
Participation rate	0.621	0.995
Hours of work (all)	16.345	41.2
Hours of work (H>0)	26.311	41.4
Gross wage rate (all)*	29.702	39.439
Gross wage rate (H> 0)	30.61	39.453
Age	37.149	39.56
High education	0.096	0.341
Low education	0.087	0.039
Net household income (per month)	7691.416	
Number of children	1.13	
Number of preschool children	0.57	
Number of schoolaged children	0.38	
Number of selected households	2450	

* Includes predicted wages for non-workers

Figure 1: Predicted and observed wage distribution

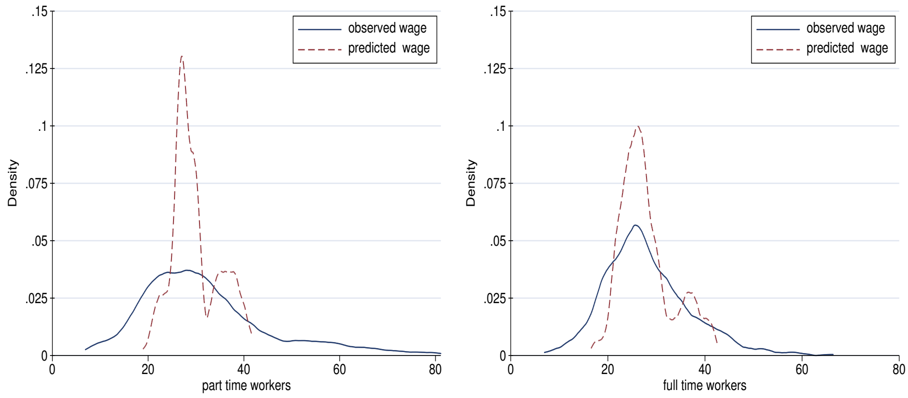

Figure 2: Distribution of female working hours

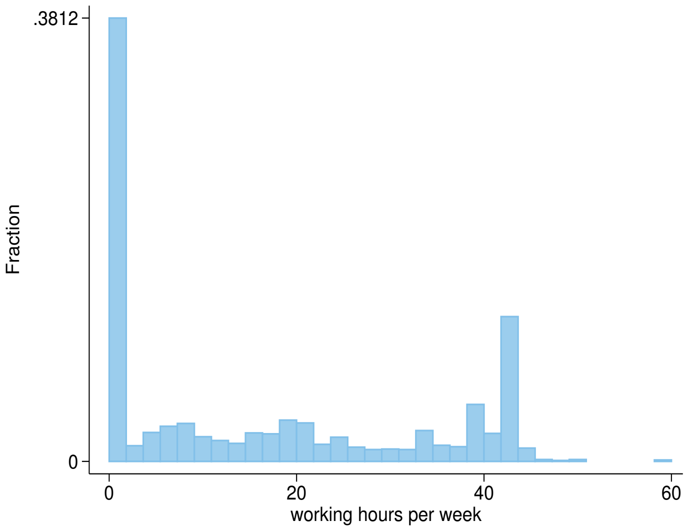

Table 2: Estimation results, model S1, S2 and S3

	S1		S2		S3	
Variable	Coeff.	(Std. Err.)	Coeff.	(Std. Err.)	Coeff.	(Std. Err.)
income^2	-0.2094	(0.1591)	-0.3461**	(0.1125)	-0.3691**	(0.1161)
hours^2	0.1667**	(0.0166)	0.0168	(0.0241)	-0.1223**	(0.0339)
hours $\times$ income	-0.1228	(0.0768)	-0.1301**	(0.0415)	-0.1984**	(0.0383)
income	3.3186**	(0.6582)	3.7072**	(0.5299)	3.7495**	(0.4996)
hours	-0.6992**	(0.1518)	0.1529	(0.1516)	1.1213**	(0.2039)
$\times$ age$-$40	-0.0324**	(0.0021)	-0.0319**	(0.0021)	-0.0362**	(0.0025)
$\times$ (age$-$40)^2	-6.7e-06	(0.0002)	2.1e-05	(0.0002)	0.0001	(0.0002)
$\times$ preschool children	-0.8241**	(0.0365)	-0.8864**	(0.0461)	-1.0724**	(0.1088)
$\times$ schoolaged children	-0.4161**	(0.0315)	-0.6008**	(0.0434)	-0.8450**	(0.0792)
$\times$ high educated	0.3958**	(0.0600)	0.4119**	(0.0622)	0.4507**	(0.0716)
fixed cost/4000			0.5678**	(0.0879)	1.0465**	(0.0833)
preschool children			-0.1277**	(0.0378)	-0.2383**	(0.0916)
schoolaged children			-0.3117**	(0.0514)	-0.4928**	(0.0660)
σ_y					2.8763**	(0.4106)
N	2450		2450		2450	
Log-likelihood	-3279.66		-3225.34		-3213.33	

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

Table 3: Estimation Results, Model SNP

Variable	Coeff.	(Std. Err.)
income^5	-0.0064	(0.0119)
$\text{income}^4 \times \text{hours}$	-0.0065	(0.0210)
$\text{income}^3 \times \text{hours}^2$	-0.0117	(0.0322)
$\text{income}^2 \times \text{hours}^3$	0.0540	(0.0414)
$\text{income} \times \text{hours}^4$	0.1499^{*}	(0.0606)
hours^5	$0.1777^{\dagger}$	(0.0930)
income^4	0.0377	(0.0810)
$\text{income}^3 \times \text{hours}$	0.0163	(0.1575)
$\text{income}^2 \times \text{hours}^2$	-0.3781	(0.3073)
$\text{income} \times \text{hours}^3$	-1.4146^{**}	(0.5335)
hours^4	-2.0008^{**}	(0.9929)
income^3	0.1029	(0.1685)
$\text{income}^2 \times \text{hours}$	0.8906	(0.6699)
$\text{income} \times \text{hours}^2$	4.3782^{**}	(1.6397)
hours^4	8.4758^{*}	(3.8700)
income^2	-0.9909	(0.6253)
hours^2	-17.0240^{*}	(6.9883)
$\text{income} \times \text{hours}$	-5.1780^{*}	(2.0893)
income	4.3300^{**}	(1.4967)
hours	16.2566^{**}	(5.9735)
$\times$ age$-$40	-0.0309^{**}	(0.0021)
$\times$ $\text{age}^2 - 40$	0.0001	(0.0002)
$\times$ preschool children	-0.9066^{**}	(0.0476)
$\times$ schoolaged children	-0.6404^{**}	(0.0499)
$\times$ high educated	0.4131^{**}	(0.0630)
fixedcost/4000	2.4307^{*}	(1.0451)
preschool children	-0.1440^{**}	(0.0384)
schoolaged children	-0.2968^{**}	(0.0541)
N	2450	
Log-likelihood	-3169.07	

Significance levels : $\dagger$: 10% $*$: 5% $**$: 1%

Table 4: Average Predicted Probabilities

choice	actual	S1	S2	S3	SNP	UC	GC
0	0.379	0.363	0.381	0.385	0.379	0.379	0.379
8	0.112	0.151	0.121	0.116	0.112	0.112	0.112
16	0.109	0.093	0.091	0.090	0.109	0.109	0.109
25	0.111	0.082	0.093	0.094	0.111	0.111	0.111
33	0.076	0.106	0.120	0.121	0.076	0.076	0.076
42	0.212	0.205	0.194	0.194	0.212	0.212	0.212
Pseudo R2		0.253	0.265	0.268	0.278	0.306	0.424

Table 5: Tests of Restrictions

mod.	log L	coeff.	vs mod.	log L	coeff.	df	LR	chi2(1%)
S2	-3225.34	13	SNP	-3169.07	28	15	112.53	30.6
S2	-3225.34	13	UC	-3044.88	137	124	360.91	164
S2	-3225.34	13	UCR	-3045.78	137	124	359.12	164
UC	-3044.88	137	GC	-2526.53	275	138	1063.7	180