On the role of unobserved preference heterogeneity in discrete choice models of labor supply

Daniele Pacifico

University of Bologna - Department of Economics, Center for the Analysis of Public Policy - University of Modena and Reggio Emilia

15. November 2009

On the role of unobserved preference heterogeneity in discrete choice models of labor supply

D. Pacifico*

Department of Economics, University of Bologna, 1 piazza Scaravilli, Bologna, 40126 Italy

Abstract

The aim of this paper is to analyse the role of unobserved preference heterogeneity in structural discrete choice models of labor supply. Within this framework, unobserved heterogeneity has been estimated either parametrically or nonparametrically through random coefficient models. Nevertheless, the estimation of such models by means of standard, gradient-based methods is often difficult, in particular if the number of random parameters is high. For this reason, the role of unobserved taste variability in empirical studies is often constrained since only a small set of coefficients is assumed to be random. However, this simplification may affect the estimated labor supply elasticities and the subsequent policy prescriptions. In this paper, we propose a new estimation method based on an EM algorithm that allows us to fully consider the effect of unobserved heterogeneity nonparametrically. Results show that labor supply elasticities and policy prescriptions do change significantly only when the full set of coefficients is assumed to be random. Moreover, we analyse the behavioural effects of the introduction of a working-tax credit scheme in the Italian tax-benefit system and show that the magnitude of labor supply reactions and the post-reform income distribution can differ significantly depending on the specification of unobserved heterogeneity.

Key words: behavioural microsimulation, labor supply, unobserved heterogeneity, random coefficient mixed models, EM algorithm

Jel Classification: J22, H31, H24, C25, C14

*Email address: d.pacifico@unibo.it

On the role of unobserved preference heterogeneity in discrete choice models of labor supply

D. Pacifico*

Department of Economics, University of Bologna, 1 piazza Scaravilli, Bologna, 40126 Italy

1. Introduction

Structural discrete choice models of labor supply are a useful tool for the *ex-ante* evaluation of labor supply reactions to tax reforms. The underlying theoretical model draws from a neoclassical environment, with optimising agents and random utility functions defined over a discrete leisure-consumption space. Both the categorisation of the leisure-consumption space and the assumption of random utilities create a typical discrete choice setting, which allows handling highly non-convex budget sets and the non-participation choice easily.

Modelling labor supply responses using a discrete approach has become increasingly popular in recent years[1]. The main idea is to simulate real consumption over a finite set of alternatives of leisure given the actual tax-benefit system. Then, under the hypothesis that agents choose the combination of leisure and consumption that maximises their random utility given the observed tax-benefit rules, the probability of the observed choice can be recovered once a (convenient) assumption on the utility stochastic term is made[2].

As for the rule of unobserved preference heterogeneity in the labor supply literature, this has mainly been considered in a parametric way by assuming that unobserved taste variability has a specific – typically continuous – distribution, which can be then integrated out from the likelihood during the estimation process. Recently, unobserved heterogeneity has been estimated nonparametrically using a latent class approach *á la* Heckman and Singer (1984). The idea is to assume a discrete distribution for the unobserved heterogeneity and to estimate the mass

*Email address: d.pacifico@unibo.it

[1]Earlier works that explore this method are those from Van Soest (1995), Keane and Moffitt (1998) and Blundell et al. (2000). See Blundell and MaCurdy (1999) for a review of alternative approaches for labor supply models.

[2]Hence, what is estimated within this framework are the parameters of the direct utility function and not of typical labor supply Marshallian functions.

points and the population shares along with the other parameters of the utility function[3].

However, regardless of the approach used, unobserved heterogeneity has always been assumed to affect only a relatively small set of parameters, in particular those that mainly define the marginal utility of consumption and/or the marginal utility of leisure. The reason for this simplification does not rest on a specific economic theory but on the computational problems that normally arise with gradient-based maximisation algorithms as Newton-Raphson or BHHH. Indeed, labor supply models contain a relatively high set of parameters so as to better explain how labor supply behaviour relates to the tax system. Moreover, the presence of random coefficients significantly changes the shape of the likelihood function, increasing its complexity and slowing down the search algorithm.

Hence, it follows that the higher the number of parameters specified as random, the more difficult and slower the numerical computation of the gradient. This implies, in turn, a more instable Hessian with the related probability of empirical singularity at some iterations. For this reason, the number of random parameters in labor supply models has always been small, which might curtails the role of unobserved heterogeneity. Thus, depending on the size of unobserved heterogeneity and on the number of coefficients specified as random, post-estimation results - as elasticities or other measures - may not differ significantly from those obtained without accounting for unobserved taste heterogeneity.

Haan (2006) proves that no matter the way the researcher accounts for unobserved heterogeneity - parametrically or nonparametrically with just a few random parameters - the subsequent labor supply elasticities do not change significantly with respect to the base model without unobserved heterogeneity. Moreover, Colombino and Locatelli (2008) compare the results of a hypothetical tax reform when unobserved heterogeneity is introduced parametrically in three coefficients and find very small differences in the evaluation of the reform. This paper confirms these previous findings although shows that a complete stochastic specification - with all the coefficients specified as random - not only improves the results in terms of fitting but also leads to highly significant differences in the subsequent labor supply elasticities. This finding is particularly important for the applied research whose aim is to evaluate the labor supply reaction to tax reforms empirically. Indeed, different elasticities of labor supply imply different policy recommendations and different judgements about the reform under analysis.

In order to estimate a fully random specification, we bypass the computational difficulties of gradient-based maximisation methods by developing a new

[3]Recent examples are from Haan (2006), Haan and Uhlendorff (2007), Wrohlich (2005), Bargain (2007) and Vermeulen et al. (2006).

Expectation-Maximisation (EM) algorithm for the nonparametric estimation of mixing distributions that is quickly implementable, ensures convergence and speeds-up the estimation process. Our empirical analysis is based on the European panel of Income and Living Conditions (EU-SILC) and is carried out in two steps. Firstly, we estimate labor supply elasticities using different specifications of unobserved taste heterogeneity and show that they can differ significantly depending on the way in which unobserved heterogeneity is specified. Secondly, we simulate a real tax reform - the introduction of a working tax-credit scheme in the Italian tax-benefit system - in order to show how different labor supply elasticities can lead to different results in terms of labor supply reactions and post-reform income distribution.

This paper is structured as follows. In section 2 we present the basic discrete choice model of labor supply. Section 3 shows how unobserved heterogeneity has been considered in the literature. Section 4 presents an overview of the EM algorithm. Section 5 comments on the estimated utility parameters and compares elasticities across various specifications of our model. Section 6 contains the simulation and the evaluation of the introduction of a UK-style working tax-credit schedule for Italy. Section 7 concludes.

2. The basic econometric model without unobserved heterogeneity

In this section we develop the econometric framework for the basic structural labor supply model. For simplicity, we focus only on married/*de facto* couples and do not consider singles. As common in this literature, we follow a unitary framework in order to model the household's decision process, which implies that the couple as a whole is the decision maker[4]. We assume that each household has a limited set of work alternatives and that spouses choose simultaneously the combination that maximises a joint utility function, which is defined over the household disposable income and the hours of work of either spouse[5]. If the household utility is subject to optimisation errors, then it is possible to recover the probability of the observed choice once an assumption on the distribution of the stochastic component is made. More formally, let $\boldsymbol{H_j} = [hf_j\,; hm_j]$ be a vector of worked hours for alternative j, hf for women and hm for men. Let y_{ij} be the net household income associated with combination j and $\boldsymbol{X_i}$ be a vector of individual and household characteristics. Then the utility of household i when $\boldsymbol{H} = \boldsymbol{H_j}$ is:

$$U_{ij} = U(y_{ij},\, \boldsymbol{H_j},\, \boldsymbol{X_i}) + \xi_{ij} \tag{1}$$

[4]See Chiappori and Ekeland (2006) for a collective model of labor supply.

[5]In a static environment, household expenditures equals household net-income. Moreover, we model the leisure decision as a work decision.

Where ξ_{ij} is a choice-specific stochastic component which is assumed to be independent across the alternatives and to follow a type-one extreme value distribution The net-household income of household i when alternative j is chosen is defined as follows:

$$y_{ij} = w_{if}hf_j + w_{im}hm_j + nly_i + TB(w_{if}; w_{im}; \boldsymbol{H}_j; nly_i; \boldsymbol{X_i}) \tag{2}$$

Where w_{if} and w_{im} are the hourly gross wages from employment for women and men respectively; nly_i is the household non-labor income and the function *TB(.)* represents the tax-benefit system, which depends on the gross wage rates, hours of work, household non-labor income and individual characteristics. It is worth noting that this function could produce highly non-linear and non-convex budget sets for most of the population of interest due to the mixing effect of tax credits, tax deductions, tax brackets and benefit entitlements[6]. Following Keane and Moffitt (1998) and Blundell et al. (2000), the observed part of the utility in eq.1 is defined as a second order polynomial with interactions between the wife and the husband terms:

$$\begin{aligned} U(y_{ij}; \boldsymbol{H}_j; \boldsymbol{X_i}) &= \alpha_1 y_{ij}^2 + \alpha_2 hf_j^2 + \alpha_3 hm_j^2 + \\ &\quad + \alpha_4 hf_j hm_j + \alpha_5 y_{ij} hf_j + \alpha_6 y_{ij} hm_j + \\ &\quad + \beta_1 y_{ij} + \beta_2 hf_j + \beta_3 hm_j \end{aligned} \tag{3}$$

In order to introduce individual characteristics in the utility function, the coefficients of the linear terms are defined as follows:

$$\beta_j = \sum_{i=1}^{K_j} \beta_{ij} x_{ij} \qquad j \epsilon \{1, 2, 3\} \tag{4}$$

Under the assumption that the couple maximises her utility and that the utility stochastic terms in each alternative are independent and identically distributed with a type-one extreme value distribution, the probability of choosing $\boldsymbol{H_j} = [hf_j; hm_j]$ is given by[7]:

$$\begin{aligned} Pr(\boldsymbol{H_j} | \boldsymbol{X_i}) &= Pr[U_{ij} > U_{is}, \forall s \neq j] \\ &= \frac{exp(U(y_{ij}, \boldsymbol{H_j}, \boldsymbol{X_i}))}{\sum_{k=1}^{K} exp(U(y_{ik}, \boldsymbol{H_k}, \boldsymbol{X_i}))} \end{aligned} \tag{5}$$

[6]For those people who are not observed working gross wage rates are estimated according with a standard selection model as in Heckman (1974). We estimated different models for either spouses and used the estimated gross wage rates for the whole sample.

[7]See McFadden (1973)

Then, the log likelihood function for the basic model is:

$$LL = \sum_{i=1}^{N} log \prod_{j=1}^{J} Pr(\boldsymbol{H_j}|\boldsymbol{X_i})^{d_{ij}} \tag{6}$$

Where d_{ij} is a dummy variable that equals to one for the observed choice and zero otherwise.

The econometric model described above is a typical conditional logit model, which can be estimated by means of high-level statistical software packages. However, the drawbacks of this basic model are well known in the literature. As pointed out in Bhat (2000) there are three main assumptions which underline the standard conditional logit specification. The first one assumes that the stochastic components of the utility function are independent across alternatives. The second assumption is that unobserved individual characteristics do not affect the response to variations in observed attributes. Finally, the assumption of error variance-covariance homogeneity implies that the extent of substitutability among alternatives is the same across individuals.

One prominent effect of these assumptions is the well-known property of *independence from irrelevant alternatives* (IIA) at an individual level, which can be very restrictive in our labor supply framework[8].

The next section introduces different models that have been used in the labor supply literature in order to reduce the extent of the IIA property by relaxing one or more of the assumptions listed above.

3. Modelling unobserved heterogeneity in preferences

The literature has developed several models that relax the IIA property of the multinomial conditional logit. Parametric random coefficients mixed models are probably the most important among numerous innovations because of their overall flexibility[9]. The idea that underlies these specifications is that agents have different unobserved tastes that affect individual response to given attributes. In other words, the parameters that enter the utility are not fixed across the population - like in traditional multinomial logit models - but vary randomly with a given unknown distribution. In empirical works, the analyst makes an assumption on the distribution of this unobserved variability and the moments of this distribution are then estimated along with the other preference parameters. Clearly, there is

[8]Consider a choice set initially defined by just two alternatives: working full time and not working. The IIA assumption implies that introducing another alternative - say a part-time alternative - does not change the relative odds between the two initial choices.

[9]See McFadden and Train (2000).

a great freedom in the choice of different densities and many alternatives can be tested[10].

However, any parametric specification has several drawbacks implied by its intrinsic characteristics. As Train (2008) points out, using a normal density, which has a support on both sides of zero, could be problematic when the unobserved taste is expected to be signed for some economic reasons (such the marginal utility of consumption). Other alternatives that avoid this problem, like the log-normal or the triangular distribution, have their own drawbacks in applied research.

Another problem of these mixed models is simply practical. Indeed, since the analyst does not observe the individual's tastes completely, the conditional probability of the observed choice has to be integrated over all possible values of the unobserved taste. Depending on the number of parameters assumed to be random, this could imply the construction of a multi-dimensional integral that becomes difficult to compute, even with simulation methods. For this reason, many researchers choose to reduce the number of random parameters so as to keep the estimation feasible, and this particularly true in the labor supply literature where the number of parameters to be estimated could be relatively high.

More formal, it is convenient to rewrite the direct utility function of equation 3 in a matrix form. In particular, let the utility of choice j for agent i be:

$$\mathbf{U}(\mathrm{y_{ij}}, \boldsymbol{H_j}, \boldsymbol{X_i}) = \boldsymbol{W}_{\mathbf{ij}}^{'}\boldsymbol{\alpha} + \boldsymbol{G}_{ij}^{'}\boldsymbol{\beta} + \xi_{ij} \tag{7}$$

With $\boldsymbol{W}_{ij} = (y_{ij}^2,\, hf_j^2,\, hm_j^2,\, hfhm_j,\, y_{ij}hf_j,\, y_{ij}hm_j)'$;$\boldsymbol{G}_{ij} = (y_{ij},\, hf_j,\, hm_j)'$ and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being the subsequent vectors of coefficients as in equation 3. Following the recent labor supply literature, assume now the set of parameters in vector $\boldsymbol{\beta}$ to be random:

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Theta}\boldsymbol{X}_i + \boldsymbol{\Omega}\boldsymbol{\vartheta}_i \quad E(\boldsymbol{\vartheta}_i) = \mathbf{0},\ Cov(\boldsymbol{\vartheta}_i) = \boldsymbol{\Sigma} \tag{8}$$

With $\boldsymbol{X}_i$ defined as the matrix of observed individual and household characteristics that affect the vector of means $\boldsymbol{\beta}$, $\boldsymbol{\Theta}$ the corresponding coefficient matrix, $\boldsymbol{\vartheta_i}$ a vector of iid unobserved individual taste shifters, $\boldsymbol{\Omega}$ the Cholesky factor of the Variance-Covariance Matrix $\boldsymbol{\Sigma}$ to be estimated along with the other structural parameters. Since $\boldsymbol{\vartheta_i}$ is not observed, the probability of the observed choice has to be integrated over its distribution. If we now let $\boldsymbol{\phi(\vartheta_i)}$ be the multivariate density of the random vector $\boldsymbol{\vartheta_i}$, the unconditional probability of choice j for household i can be now written as:

$$Pr(\boldsymbol{H_{ij}}|\boldsymbol{X_i}) = \int Pr(\boldsymbol{H_{ij}}|\boldsymbol{X_i}, \boldsymbol{\vartheta_i})\boldsymbol{\phi(\vartheta_i)}d\boldsymbol{\vartheta_i} \tag{9}$$

[10]Common choices are the Gaussian, the log-normal or the triangular distribution.

Where $Pr(H_i = H_{ij} \,|\, \boldsymbol{X_i}, \boldsymbol{\vartheta_i})$ is the conditional logit probability of choice j as defined in equation 5. Since this multidimensional integral cannot be solved numerically, Train (2003) suggests simulation methods with Halton sequences. The simulated-log likelihood for the sample is then:

$$LL = \sum_{i=1}^{N} log \frac{1}{R} \sum_{r=1}^{R} \prod_{j=1}^{J} Pr(H_{ij} | \boldsymbol{X_i}, \boldsymbol{\vartheta_{ir}})^{d_{ij}} \tag{10}$$

Where the integrals are approximated by the empirical expectation over the R draws from the selected multivariate distribution of the unobserved tastes. The literature has recently suggested latent class logit models as a variant of the standard multinomial logit that resembles the random coefficients mixed model described above. Latent class models can account for unobserved heterogeneity nonparametrically and have been proposed so as not to be constrained by distributional assumptions. These models were developed theoretically in the eighties by Heckman and Singer (1984) and have received great attention in the area of models for count. First applications of this method to discrete choices models are those in Swait (1994) and Bhat (1997). The idea behind these models is that agents are sorted in a given number of classes and that agents who are in different classes have different preference parameters and hence different responses to given attributes. The analyst does not observe the class membership and needs to model the probability of class membership along with the probability of the observed choice. Let us assume that there are C latent classes in the population of interest. As for the previous mixed model, we follow the recent labor supply literature and assume that only the preference parameters in vector $\boldsymbol{\beta}$ of equation 7 differ among people in different classes. Later, we will generalise our model and assume that the whole set of taste parameters differs among classes. The conditional logit probability that household i belonging to class c chooses alternative j is:

$$Pr(\boldsymbol{H_{ij}} | \boldsymbol{X_i}, \boldsymbol{\beta_c}) = \frac{exp(\boldsymbol{W}'_{ij}\boldsymbol{\alpha} + \boldsymbol{G}'_{ij}\boldsymbol{\beta_c})}{\sum_{k=1}^{K} exp(\boldsymbol{W}'_{ik}\boldsymbol{\alpha} + \boldsymbol{G}'_{ik}\boldsymbol{\beta_c})} \tag{11}$$

Since class membership is not observed, the analyst has also to model the probability for each household to belong from each latent class. Following the latent class literature, we adopt a multinomial logit formula in order to keep these unconditional probabilities in their right range and to ensure that they sum up to one for every household[11]:

$$Pr(class_i = c|\boldsymbol{\Delta_i}) = \frac{exp(\boldsymbol{\Delta_i'}\boldsymbol{\gamma_c})}{\sum_{c=1}^{C} exp(\boldsymbol{\Delta_i'}\boldsymbol{\gamma_c})}, c = 1, .., C; \boldsymbol{\gamma_C} = \mathbf{0} \tag{12}$$

Where $\boldsymbol{\gamma_c}$ is a vector of unknown class parameters that specifies the contribution of the observed individual characteristics contained in the matrix $\boldsymbol{\Delta_i}$ to the probability of latent class membership[12].

As Roeder et al. (1999) point out, the variables in matrix $\boldsymbol{\Delta_i}$, which are traditionally called *risk factors*, have to be specified properly. Nevertheless, in many applications, and in particular those related to the labor supply literature, they normally collapse to just a simple scalar in order to simplify the analysis and to speed-up estimation.

Given equations 11 and 12, the conditional probability that a randomly selected household i chooses alternative j is:

$$\sum_{c=1}^{C} Pr(class_i = c \,|\, \boldsymbol{\Delta_i}) Pr(\boldsymbol{H_{ij}}|\boldsymbol{X_i},\boldsymbol{\beta_c}) \tag{13}$$

Hence, the log-likelihood for the whole sample is:

$$LL = \sum_{i=1}^{N} log \sum_{c=1}^{C} Pr(class_i = c|\boldsymbol{\Delta_i}) \prod_{j=1}^{J} Pr(\boldsymbol{H_{ij}}|\boldsymbol{X_i},\boldsymbol{\beta_c})^{d_{ij}} \tag{14}$$

As Train (2008) points out, differently from parametric random coefficients mixed models, the primary difficulty with this nonparametric approach is computational rather than conceptual since standard gradient-based algorithms for maximum likelihood estimation become increasingly difficult when the number of latent classes rises.

Importantly, these empirical difficulties, which closely resembles those encountered in the parametric mixed model described above, explain why labor supply analysts significantly constrain the number of latent classes, the number of risk factors and the number of parameters that can differ in each class[13].

[11] See Greene (2001).

[12] The C*th* vector of parameters is normalised to zero to ensure identification.

[13] Interestingly, as we have seen with the two mixed models, the set of parameters that are traditionally assumed to be random in the labor supply literature (i.e. the parameters in vector $\boldsymbol{\beta}$, according to our specification) are the same whether the analysis is carried out parametrically with continuous random coefficients mixed logit models or nonparametrically with latent class models.

To summarise, the two mixed models outlined so far share a similar computational problem, which largely depends on the algorithms that are traditionally used for the estimation of such models.

Mainly due to these difficulties, the role of unobserved heterogeneity in the labor supply literature has always been limited and this could partially justify Haan's claim, who has not found significant differences in the labor supply elasticities obtained when unobserved heterogeneity is introduced parametrically or nonparametrically. We indeed confirm Haan's findings in our empirical analysis although we show that when unobserved heterogeneity is considered in a more comprehensive way, the subsequent labor supply elasticities do change significantly.

Precisely, our intuition is to develop a new estimation method that is not completely based on a standard gradient-based optimisation process so that the computational difficulties outlined in this section can be avoided. In particular, following Train (2008), we propose an EM algorithm for the nonparametric estimation of mixing distributions that, given its overall stability, does ensure convergence and speeds-up the computational process. Therefore, we can explore the role of unobserved heterogeneity in a very general way since we are constrained neither to distributional assumptions nor to computational difficulties.

4. An EM recursion for discrete choice models of labor supply

EM algorithms were initially introduced to deal with missing data problems, although they turned out to be a very good method of estimating latent class models where the missing data is the class shares[14]. Nowadays, they are widely used in many economic fields where the assumption that people can be grouped in classes with different unobserved taste heterogeneity is reasonable. Hence, many applications of this recursion can be found in health economics or consumer-choice modelling but, as long as we know, there is no evidence for labour supply models.

From an econometric point of view, the attractiveness of this estimation method lies in its overall stability. Moreover, Train (2008) has shown how EM algorithms can be used for the nonparametric estimation of mixing distributions.

The recursion is known as "E-M" because it consists of two steps, namely an "Expectation" and a "Maximization". The term being maximized is the expectation of the *joint* log-likelihood of the observed and missing data, where this expectation is over the distribution of the missing data conditional on the density of the observed data and the previous parameters estimates. Consider the latent class model outlined in the previous section. Traditionally, the log-likelihood in eq.14 is maximized by standard gradient-based methods as Newton Raphson or

[14]Our EM recursion is partially based on the algorithm developed in Train (2008). The routine is coded in STATA 10 and is freely available in Pacifico (2009).

BHHH. However, it can be shown that the same log-likelihood can be maximized by repeatedly updating the following recursion:

$$\boldsymbol{\eta}^{s+1} \quad = \quad argmax_{\eta} \sum_i \sum_c C_i(\boldsymbol{\eta}^s) ln(L_i|class_i = c) \tag{15}$$

Where $L_i|class_i = c$ is the *missing-data* log-likelihood, which is defined by the product of the unconditional density of the missing data $w_{ic}(\boldsymbol{\gamma_c}) = \frac{exp(\boldsymbol{\Delta}_{\boldsymbol{i}}^{\prime}\gamma_c)}{\sum_{c=1}^{C} exp(\boldsymbol{\Delta}_{\boldsymbol{i}}^{\prime}\gamma_c)}$ (as in eq.12) and the density of the observed choice: $\prod_j P(\boldsymbol{H_{ij}}|\boldsymbol{X_i}, \boldsymbol{\pi_c})^{d_{ij}}$, $\boldsymbol{\pi_c} = (\boldsymbol{\beta_c}; \boldsymbol{\alpha_c})'$, $\boldsymbol{\eta} = (\boldsymbol{\pi_c}; \boldsymbol{\gamma_c}, c = 1, 2, .., C)$ and $C(\boldsymbol{\eta}^s)$ is the posterior probability that household i belongs to class c, conditional on the density of the observed choice and the previous value of the parameters. This conditional probability, $C(\boldsymbol{\eta}^s)$, is the key future of the EM recursion and can be computed by means of Bayes' theorem:

$$C_i(\boldsymbol{\eta}^s) = \frac{L_i|class_i = c}{\sum_{c=1}^{C} L_i|class_i = c} \tag{16}$$

Now, given that:

$$ln\, w_c(\boldsymbol{\gamma_c}) P(\boldsymbol{H_{ij}}|\boldsymbol{X_i}, \boldsymbol{\pi_c}) = ln\, w_c(\boldsymbol{\gamma_c}) + ln\, P(\boldsymbol{H_{ij}}|\boldsymbol{X_i}, \boldsymbol{\pi_c}) \tag{17}$$

the recursion in eq.15 can be split into different steps:

1. Form the contribution to the likelihood $(L_i \,|\, class_i = c)$ as defined in eq.15 *for each class*[15],
2. Form the *individual-specific* posterior probabilities of class membership using eq.16,
3. *For each class*, maximise the *weighted* log-likelihood so as to get a new set of $\pi_c,\ c = 1, ..., C$:

$$\pi_c^{s+1} = argmax_{\boldsymbol{\pi}} \sum_i C(\boldsymbol{\eta}^s) ln \prod_j P(\boldsymbol{H_{ij}}|\boldsymbol{X_i}, \boldsymbol{\pi_c})^{d_{ij}} \tag{18}$$

4. Following eq.17, maximise the other part of the log-likelihood in eq.14 and get a new set of $w_c, c = 1, 2, ..., C$:

$$w_{ic}^{s+1} = argmax_{\boldsymbol{w}} \sum_{i=1}^{N} \sum_{c=1}^{C} C_i(\boldsymbol{\eta}^s) ln\, w_{ic}(\boldsymbol{\gamma_c}) \tag{19}$$

[15]For the first iteration, starting values have to be used for the densities that enter the model. Importantly, these starting values must be different in every class otherwise the recursion estimates the same set of parameters for all the latent classes.

(a) In particular, compute the new parameters that specify the impact of the risk factors as:

$$\boldsymbol{\gamma}^{s+1} = argmax_{\boldsymbol{\gamma}} \sum_{i=1}^{N} \sum_{c=1}^{C} C_i(\boldsymbol{\eta}^s) ln \frac{exp(\boldsymbol{\Delta}_{\boldsymbol{i}}^{\boldsymbol{\prime}} \boldsymbol{\gamma_c})}{\sum_c exp(\boldsymbol{\Delta}_{\boldsymbol{i}}^{\boldsymbol{\prime}} \boldsymbol{\gamma_c})} \tag{20}$$

Where $\boldsymbol{\gamma_C} = \mathbf{0}$ for identification

(b) And then update $w_{ic}(\boldsymbol{\gamma_c})$, $c = 1, ..., C$ as:

$$w_{ic}^{s+1} = \frac{exp(\boldsymbol{\Delta}_{\boldsymbol{i}}^{\boldsymbol{\prime}} \boldsymbol{\hat{\gamma}_c^{s+1}})}{\sum_c exp(\boldsymbol{\Delta}_{\boldsymbol{i}}^{\boldsymbol{\prime}} \boldsymbol{\hat{\gamma}_c^{s+1}})}, c = 1, 2, ..., C; \boldsymbol{\gamma_C} = \mathbf{0} \tag{21}$$

5. Once π_c^s, γ^s and w_c^s have been updated to iteration *s+1*, the posterior probability of class membership $C(\boldsymbol{\eta}^{s+1})$ can also be recomputed and the recursion can start again from point 3 until convergence[16].

It is worth noting that in each maximization, the posterior probability of class membership enters the log-likelihood without unknown parameters to be estimated and can be seen as an individual weight. Hence, eq.18 defines a typical conditional logit model with *weighed observations* that can be estimated easily with respect to the maximization of the whole model as in eq.14.

Importantly, the EM algorithm has been proved to be very stable and, under conditions given by Dempster et al. (1977) and Wu (1983), this recursion always climbs uphill until convergence to a local maximum[17].

With this model in hand, it is possible to estimate a full latent class model of labor supply without being conditioned neither to the number of parameters assumed to be random nor to the number of latent classes. Moreover, the estimation time drops significantly with respect to the time spent by standard gradient-based algorithms used for the estimation of the other models[18].

[16] Train (2008) does not use demographics for the class shares. In this case point 4 is replaced with:

$$w_c^{s+1} = \frac{\sum_i C_i(\boldsymbol{\eta}^{s+1})}{\sum_i \sum_c C_i(\boldsymbol{\eta}^{s+1})}, c = 1, ..., C \tag{22}$$

Where $C_i(\boldsymbol{\eta}^{s+1})$ is computed using the updated values of π_c (from point 3) and the previous values of the class shares.

[17] Clearly, it is always advisable to check whether the local maximum is also global by using different starting values.

[18] Both the continuous random coefficient mixed logit models and the latent class model *á la* Heckman and Singer (1984) are very time consuming when estimated via maximum likelihood. With about 30 parameters and 4000 observations, the STATA routines take about 6 hours to get

5. Empirical findings

For our empirical analysis we use the 2006 Italian wave of the European Union panel on Income and Living Conditions. We focus on the main category of tax-payer, i.e. households of employed, and allow for a flexible labor supply for both spouses. Drawing on previous literature, all couples in which either spouse is elder than 65, self-employed, student, retired or serving in the army are excluded.

The sample selection leads to about 4000 households, which are representative of almost 60% of Italian tax-payers. The number of working hours of both women and men is categorized according to their empirical distributions. In particular, we define 6 categories of hours for women (no work, 3 part-time options and 2 full-time alternatives) and 3 for men (no work, full-time and overwork), which implies 18 different combinations for each household[19]. The disposable net household income for each alternative is derived on the basis of a highly detailed tax-benefit simulator - MAPP06 - developed at the Centre for the Analysis of Public Policies (CAPP)[20].

In table 1 we report the estimated coefficients of the three models introduced in sections 2 and 3. The first model is estimated without accounting for unobserved heterogeneity and is then a typical multinomial conditional logit (MNL) as explained in section 2.

The second model is by far the most common in the applied labor supply literature and it is normally referred to as the continuous random coefficients mixed logit (RCML), which allows for unobserved heterogeneity using a parametric assumption for its distribution. In particular, following the traditional labor supply modelling, we allow the three coefficients of the linear terms of the utility to be random with independent normal densities[21]. We then estimate the means and the standard deviations of these coefficients along with the other preference parameters using Simulated Maximum Likelihood[22].

The third model we present is the nonparametric version of the previous one, meaning that we allow the same subset of coefficients to be random and estimate them using a latent class specification. This manner of accounting for unobserved heterogeneity is becoming widespread and is commonly defined as a nonparametric estimation of mixed logit models *á la* Heckman-Singer (HSML). The model is

---

convergence with our Intel quad-core PC with 4GBs of RAM (and STATA 10.1 MP); instead, our EM recursion takes less than 1 hour to get convergence for a model with 4 latent classes and 115 parameters.

[19] The categories for women are: 0, 13, 22, 30, 36 and 42 weekly hours of work. For men we define 3 categories: 0, 43 and 50 weekly hours of work.

[20] See Baldini and Ciani (2009).

[21] The estimation with correlated normal densities did not improve the likelihood and the estimated correlation coefficients were not significant.

[22] See Train (2003).

estimated via Maximum Likelihood and for each random parameter we estimate its mass points and its population shares. As in any latent class analysis, a primary goal is the definition of the proper number of latent classes. However, as we explained in section 3, due to the computational difficulties related to standard optimization methods, labor supply analysts tend to specify a very small number of latent classes and do not include covariates in the set of risk factors. We then follow this standard specification and estimate a model with just 2 latent classes and only a constant in the set of variables that enter the probability of class membership[23].

[table 1: about here]

As results in table 1 show, most coefficients have the expected sign over the three specifications[24]. Following Van Soest (1995), we computed the first and the second derivative of the utility function with respect to income and spouses' hours of work in order to check if the empirical model is coherent with the economic theory. Results show that the marginal utility of income increases at a decreasing rate for all the households in the sample and this result holds over the three specifications[25].

If we now observe the maximized log-likelihood, we can deduce that unobserved heterogeneity is actually present in our sample. Both the models that account for unobserved taste variability dominate the simple conditional logit model. In particular, the standard deviations of the random terms in the RCML are significantly different from zero, meaning that there is a high dispersion in the utility of income and (dis)utility of work due to unobserved tastes. Importantly, the same conclusion can be derived from the HSML model where the probability of each latent class and the various mass points are highly significant. Since the two models are not nested, we use the Bayesian Information Criteria and conclude that the latent class specification dominates the RCML model. This implies that unobserved heterogeneity could be better considered in a nonparametric way.

These three different specifications are what the literature has suggested so far. As underlined before, the main problems with the RCML and the HSML

[23] Actually, we tried to estimate more sophisticated versions of the HSML model. In particular, we tried to rise the number of latent classes and to allow for covariates in the set of risk factors. Nevertheless, the estimation of any of these versions via maximum likelihood did not achieve convergence.

[24] An economic interpretation of the various coefficients is omitted here because this is not the aim of this paper. However, Baldini and Pacifico (2009) discuss and analyse widely a similar model for the Italian case.

[25] In the MLN, the marginal utility of work is negative for almost 75% of the women and for about 55% of men. Similar results are found for the other two specifications.

are both conceptual and computational. Thus, convergence and speediness are achieved at the cost of reducing the role of unobserved heterogeneity so that only few coefficients are allowed to be random.

We now present the estimates for our fourth model, which generalizes the HSML model by defining a complete latent class mixed logit specification (LCML). For the estimation of such a model, traditional gradient-based methods are still feasible but, depending on the number of latent classes, they could be highly time-consuming and could not guarantee convergences[26]. Hence, the LCML is estimated throughout the EM recursion outlined in the previous section, which allows for a great flexibility in the selection of the number of latent classes. Following Greene and Hensher (2003) and Train (2008), we adopt the Bayesian Information Criteria for the selection of the right number of latent classes. As we can see from table A-1 in the appendix, the appropriate number of latent classes according to the BIC is four.

Another important issue that the EM algorithms enable us to consider properly without computational constraints is the right specification of the "risk factors" that enter the probability of belonging to a given class. In order to account for as much information as possible in the definition of these variables, we performed a principal-component factor analysis of the correlation matrix of a set of covariates thought to be helpful for the explanation of class memberships. Table A-2 in the appendix shows the (rotated) factor loadings obtained with the varimax rotation whose eigenvalues were higher than one[27]. Following **?**, the households' risk factors that enter the probability model outlined above are then computed by using the scoring coefficients obtained through a standard regression model.

Table 2 reports the coefficients for the LCML model with four latent classes along with their (weighted) average across the four classes[28]. As can be seen, the maximized log-likelihood is significantly higher with respect to the other models and also the fitting significantly increases[29]. Looking at the sign (and magnitude) of the average coefficients, we can see that the economic implications related to this model are in line with those from the other specifications. Importantly, using the

[26] We tried to estimate this specification by ML. However, this was feasible only for the model with two latent classes since no convergence was achieved for models with a higher number of classes. Moreover, the estimation took more than 13 hours with the PC described in footnote 18.

[27] As can be seen from the magnitude of the factor loadings, the first principal factor is linked to the socio-demographic characteristics, the second and the third are related to the wife's and the husband's health conditions respectively whilst the last captures the socio-economic status.

[28] Standard errors are estimated by nonparametric bootstrap. For the bootstrap exercise we used 50 bootstrap samples, each of them having the same size of the original sample.

[29] Table A-3 in the appendix shows the predicted and actual frequencies for each alternative over our four specifications.

estimated posterior probability of class membership, it is possible to disentangle the type of households that are more representative in each class. In particular, class 1 is mainly composed of households living in southern Italy, with young children and with relatively young parents. Class 3, instead, is composed mainly by the same type of households but living in northern Italy. Interestingly, these households have, on average, a higher education then those in class 1 and are more likely to own their house. Class 4, in comparison, mainly consists of relatively older households, with less young children and with relatively worse parents' health conditions. As for the analysis of preferences in each class, we computed the marginal (dis)utility of income (work) in every class and evaluated the results using the probabilities of class membership. Interestingly, on average, households that are more likely to belong to class 1 and 3 have the lowest marginal utility of income, which could be partially explained by the relatively young age of both parents. Moreover, households with a highest probability to belong to class 1 - which are mainly located in southern Italy - have a higher marginal disutility of work if compared with the other classes[30].

[table 2: about here]

We now turn to the main issue of this paper and compute the (average) elasticities across the various specifications of our labor supply models. Following Creedy and Kalb (2005), we computed such elasticities numerically. It is worth noting that these elasticities have to be interpreted carefully because they can depend substantially on the initial discrete hour level and the relative change in the gross hourly wages. However, they are surely a useful measure of the labor supply behaviour implied in our estimated model and can be used to check whether different specifications lead to different policy prescriptions[31].

Labor supply elasticities are computed for each spouse as follows. Firstly, gross hourly wages are increased by 1% for either spouse and a new vector of net household income for each alternative is computed. Secondly, the probability of each alternative is evaluated for both the old and the new vector of net household income according to the various specifications of our model. Thereafter, the expected

[30] Many other analysis about the characteristics of households in different latent classes could be made but we defer them to other - more applied - studies.

[31] Indeed, different elasticities across the various specifications would imply different labor supply reactions to tax reforms. This, in turns, implies different results in terms of social welfare evaluation, government expected expenditure/savings and expected changes in the post-reform distribution of income.

labor supply can be computed for each household as:

$$E[H^s \,|\, Y_p^s, \boldsymbol{X_i}] = \sum_{k=1}^{K^s} Pr(H_k^s \,|\, Y_p^s, \boldsymbol{X_i}) \cdot hours_k^s$$

Where s=*men, women* and p=*after, before*. Finally, the labor supply elasticities for either spouse are defined as:

$$\varepsilon_s = \frac{E[H^s \,|\, Y_{after}^s, \boldsymbol{X_i}] - E[H^s \,|\, Y_{before}^s, \boldsymbol{X_i}]}{E[H^s \,|\, Y_{before}^s, \boldsymbol{X_i}]} \cdot \frac{1}{0.01}$$

In order to check whether different specifications lead to different labor supply elasticities, we adopt the same strategy as Haan (2006). More specifically, we computed 95% bootstrapped confidence intervals for the MNL labor supply elasticities and checked whether they differ significantly from those obtained with other specifications. Table 3 shows the (average) own elasticities derived from 1% increase in the gross hourly wages of either spouse. As can be observed, women's elasticities are higher than men's elasticities. Female cross elasticities are not significantly different from zero whilst male cross elasticities are relatively higher and positive. If we now look at the elasticities divided by socio-demographic characteristics, we can see that elasticities are higher in the case of households in southern Italy (which is the poorest part of the country) and for people with lower education. Children reduce labor supply elasticities in particular if they are either many or young. These findings are common across the various specifications although the magnitude is always slightly bigger for those models that account for unobserved heterogeneity. Importantly, the parametric random coefficient mixed logit and the latent class model with only few random coefficients produce very similar results in terms of estimated elasticities. Moreover, as found also in Haan (2006), these elasticities always fall inside the 95% confidence interval for the elasticities derived from the conditional logit model. However, if we now consider the elasticities produced with the LCML model, they are significantly higher and always fall outside the confidence intervals constructed for the MNL specification, meaning that we cannot reject the hypothesis of different values.

[table 3: about here]

These findings are relevant in particular for the applied literature. Indeed, discrete choice labor supply models have been estimated only using the RCML or the HSML so far and the estimated coefficients are then used to analyse the labor supply behaviour after specific proposals of tax reforms. However, we have shown that if unobserved heterogeneity is considered in a more comprehensive way, the resulting elasticities might be significantly different, which in turn may imply different conclusions in the subsequent welfare and distributive analysis, with the probability of suggesting different policy prescriptions related to a specific tax reforms.

In order to prove this last claim, we evaluate a real structural reform of the Italian tax-benefit system in the next section. In particular, we analyse the labor supply reaction to the introduction of a UK-style working tax credit in the Italian tax-benefit system and show that income distribution and labor supply implications are significantly different depending on the approach used.

6. Simulating a WTC for Italy

The aim of working-tax credits is to encourage the participation of low income households in the labor market. In particular, this in-work support is conditional on either of the spouses in the family working at least h hours per week and eligibility is based on gross household income. The maximum amount of this benefit is defined according to a series of individual characteristics such as number of young children, age, actual number of worked hours and presence of disability. Normally, given eligibility and the maximum payable amount, the actual benefit is a decreasing function of gross household income after a given income threshold.

Our simulation closely replicates the eligibility criteria and the main elements of the UK WFTK[32]. In particular, our WTC is composed of five elements. A basic element of €1000 for those people who are eligible; a "partner element" of €600 in case of married/*de facto* couple; a "+50" element of €100 if the person starts working after a period of inactivity and he/she is over 50 years old; a "disability element" whose amount depends on the level of certified disability (€400 for low disability + €200 in case of high disability); a child element that depends on the number and the age of children (for each child less than 3 years old the family gets €600 and for children between 3 and 6 years old eligible families get €200 per child); a "+36 element" of €300 if the person works more than 36 hours per week.

The maximum payable amount is given by the sum of these elements. Given eligibility, the effective amount paid depends on the gross household income. In particular, according to the US version of the working tax credit - the EITC - our benefit first increases until it reaches its maximum amount at the household income threshold of €16000 and then it starts decreasing sharply until zero between €16000 and €21000. As in the UK-version, eligibility depends on age, disability level and number of worked hours per week. In particular, people younger than

[32]See www.direct.gov.uk for more details.

25 years old who work at least 16 hours per week can get the benefit either if they have young children or if they have a certified level of disability. Otherwise, only people over 25 years who work for at least 30 hours are eligible. For married/*de-facto* couples, the benefit is primarily computed on an individual basis and the actual amount paid is the highest among the two spouses. In our simulation we do not enforce tax neutrality and assume that the reform is financed through new government expenditures. Grossing up our results for the selected sample of households, we predict an increment of public spending of 2.8 billion of euro for Italian married couples.

In what follows, we study the effect of this tax reform on household labor supply. Given the intrinsic probabilistic nature of our model, we aggregate the (household) probability of choosing a particular alternative of working hours so as to obtain individual frequencies for the main categories of working time. In particular, for women, we aggregate the household probability so as to get the individual frequencies of non-participation, part-time work (16-30) and full-time work (>30). For men, we only distinguish between participation and full-time work. Table 4 shows these aggregate frequencies before and after the reform for each specification of our model.

As it can be seen, the sign of the labor supply reaction is the same in all four specifications of our model. In particular, all models predict positive participation incentives for married women whilst we observe a small participation disincentive for men. Looking at the intensive margin, the highest incentive for those women who would like to participate in the labor market is for full-time jobs, although there are also positive incentive for part-time options.

If we now turn to the differences among the four models, it could be seen that the MNL, the RCML and the HSML share a very similar labor supply pattern after the reform. However, according to the elasticities computed in the previous section, the labor supply reaction produced by the LCML model is significantly stronger with respect to the other specifications.

[table 4: about here]

In order to better understand the differences between the four models, in graph 1 we report, for each decile of gross household income, the absolute difference in the average frequencies of each labor supply category before and after the reform. As expected, mainly households in the lowest decile change their labor supply behaviour. However, the overall pattern of labor incentives is quite different if we consider the LCML model with respect to the other three specifications, which share a very similar pattern across the various decile.

If we focus on the latter specifications we can see that the participation rates of married women increase the most for the second, third and fourth decile whilst the part-time incentives are stronger and positive mainly for those women from the middle class although negative for women in the first and second decile. Finally, the full-time incentives are stronger for women in the first and second decile.

If we now consider the same work incentives using the LCML specification, we observe first a significant different magnitude and, second also a different structure of incentives across the various decile, in particular for the first two. To be precise, the participation rates strongly increase for women in the first and second decile whilst part-time incentives are always positive.

The participation rates for men decrease in the four models, although the LCML model produces, again, a stronger reaction, in particular for low-income households.

[graph 1: about here]

In order to evaluate how the income distribution changes after the reform, we compute the Gini index before and after the introduction of the WTC. As it can be seen in table 5, the starting level of inequality is almost 32.3%. However, after the reform, income inequality slightly reduces. However, the results for the LCML are - again - stronger, implying a higher reduction in income inequality (-1.2% versus an average of -0.84 over the other three specifications).

[table 5: about here]

7. Conclusions

The aim of this paper has been twofold. First, we have shown that the way researchers account for unobserved heterogeneity can have an impact on the derived labor supply elasticities, which in turn implies that policy recommendations related to given tax-reforms can change significantly according to the specification of the model.

In particular, we have computed average elasticities for either spouses and proved that these elasticities could differ significantly depending on the way unobserved heterogeneity is considered. Then, we simulated a structural tax reform by introducing a working tax credit schedule in the Italian tax-benefit system and shown that policy implications, again, depend on the specification of unobserved heterogeneity.

Second, we have provided a relatively plain alternative to fully consider the effect of unobserved heterogeneity nonparametrically. In particular, we have proposed an easily-implementable EM algorithm that allows us to increase the number of random coefficients in the specification, ensure convergence and speed-up the estimation process with respect to other standard gradient-based maximization algorithms.

Acknowledgement

I would like to thank Kenneth Train for his useful suggestions about the implementation of the EM algorithm described in this paper. I am also grateful to Massimo Baldini for his important comments and for providing me the codes of the tax-benefit microsimulation model MAPP06.

Appendix

Table A-1 Latent class models with different number of classes

Latent Classes	Log-Likelihood	Parameters	BIC
1	-8069.31	25	16138.62
2	-7859.82	55	15917.76
3	-7781.35	85	15868.88
4	**-7691.49**	**115**	**15797.22**
5	-7637.51	145	15797.32

Table A-2 Rotated factor loadings

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Number of children <16	**-0.70**	0.06	-0.06	0.02
Youngest child 0-6	**-0.77**	0.04	-0.01	0.07
Southern Italy	0.00	0.16	-0.12	**-0.45**
Husband's education	-0.06	0.08	0.05	**0.78**
Wife's education	-0.19	0.08	0.04	**0.78**
House ownership	0.3	0.02	-0.03	**0.45**
Wife's age	**0.87**	-0.09	-0.13	-0.04
Husband's age	**0.86**	-0.08	-0.15	-0.09
Wife's health status	0.22	**-0.7**	-0.26	-0.1
Husband's health status	0.22	-0.23	**-0.71**	-0.12
Wife's chronic diseases	-0.02	**0.8**	0.03	-0.05
Husband's chronic diseases	-0.04	0.09	**0.77**	-0.09

Table A-3 Observed and predicted frequencies

Alternative	hours women	hours men	Observed	LCLM	MNL	RCML	HSML
1	0	0	5.76%	5.78%	5.76%	5.69%	5.73%
2	0	43	32.88%	32.88%	33.08%	33.22%	33.18%
3	0	50	12.21%	12.15%	12.01%	11.90%	11.95%
4	13	0	0.13%	0.11%	0.08%	0.07%	0.07%
5	13	43	2.44%	2.51%	3.25%	3.26%	3.26%
6	13	50	0.91%	1.03%	1.09%	1.09%	1.10%
7	22	0	0.38%	0.44%	0.25%	0.24%	0.24%
8	22	43	7.36%	6.97%	4.95%	4.96%	4.95%
9	22	50	2.34%	2.37%	1.66%	1.68%	1.68%
10	30	0	0.28%	0.29%	0.50%	0.51%	0.51%
11	30	43	3.88%	4.12%	6.74%	6.70%	6.69%
12	30	50	1.65%	1.40%	2.28%	2.30%	2.29%
13	36	0	0.76%	0.52%	0.74%	0.78%	0.77%
14	36	43	10.66%	10.68%	8.75%	8.71%	8.71%
15	36	50	2.23%	2.77%	2.89%	2.93%	2.91%
16	42	0	1.07%	1.19%	1.04%	1.10%	1.09%
17	42	43	10.87%	10.92%	11.31%	11.23%	11.25%
18	42	50	4.19%	3.86%	3.60%	3.64%	3.61%

Note: our computation based on the selected sample from EU-SILC (2006)

List of tables and graphs

Table 1 Estimated utility parameters (1)

		Coef	z	Coef	z	Coef	z
α_1:	Constant	-30.04	-7.36	-36.64	-7.81	-35.54	-7.72
α_2:	Constant	-0.08	-2.80	-0.09	-2.96	-0.09	-2.93
α_3:	Constant	-0.22	-13.94	-0.36	-8.26	-0.31	-11.00
α_4:	Constant	-2.02	-7.48	-2.18	-7.05	-2.36	-6.92
α_5:	Constant	2.38	6.14	2.76	6.31	2.65	6.15
α_6:	Constant	2.49	5.97	2.86	5.51	2.67	5.39
β_1:	Constant	50.98	19.56	61.67	17.85	-	-
	Wife's age	0.81	1.12	2.14	1.85	1.56	1.86
	Husband's age	-2.01	-3.15	-1.92	-2.88	-1.97	-2.87
	Youngest child 0-6	-7.17	-3.00	-8.12	-3.08	-9.18	-3.51
	σ_1	-	-	0.06	3.01	-	
β_2:	Constant	-0.58	-2.75	-0.89	-3.96	-	-
	Wife's age	0.06	0.48	0.0003	0.02	0.04	0.34
	Wife's age^2	-0.03	-2.46	-0.04	-2.62	-0.04	-2.76
	Wife's education	-0.21	-6.91	-0.3	-8.47	-0.30	-8.54
	Southern Italy	-0.19	-7.29	-0.18	-6.92	-0.19	-7.10
	Youngest child 0-6	0.2	2.05	0.25	2.27	0.29	2.65
	Numb. of children	-0.16	-5.36	-0.16	-5.21	-0.16	-5.16
	σ_2	-	-	0.02	1.82	-	-
β_3:	Constant	-1.3	-8.23	-0.59	-1.90	-	-
	Husband's age	0.05	0.39	0.55	2.05	0.62	2.49
	Husband's age^2	-0.01	-1.04	-0.09	-2.83	-0.09	-3.27
	Husband's educ.	-0.13	-3.72	-0.06	-1.05	-0.08	-1.70
	Southern Italy	-0.08	-2.63	-0.23	-3.68	-0.23	-4.41
	Youngest child 0-6	0.24	2.10	0.27	2.00	0.32	2.48
	σ_3	-	-	0.75	6.12	-	-
1(husb=0 ho.):	Constant	-3.14	-10.07	-3.67	-10.81	-3.53	-10.64
1(wife=0 ho.):	Constant	3.72	14.40	3.79	14.62	3.80	14.65
β_1:	Mass 1					59.5	13.4
β_1:	Mass 2					63.31	17.11
β_2:	Mass 1					-0.83	-3.13
β_2:	Mass 2					-0.80	-3.45
β_3:	Mass 1					-1.73	-6.75
β_3:	Mass 2					-0.70	-2.61
	prob. (class1)					0.78	5.18
	Log-Likelihood:	-8069		-8050		-8043	

Note: RCLM estimated by SML with 500 halton draws; the σs are the estimated standard deviations in the RCLM specification. The logit probability of class 1 is estimated for the HS model, the standard error reported in the table is computed using the "delta method". 1(husb=0 ho.) is a dummy that is equal to one for the alternatives where the husband does not work; 1(wife=0 ho.) is the same for the wife.

Table 2 Estimated utility parameters (2)

	lc. 1	z	lc. 2	z	lc. 3	z	lc.4	z	Aver.	z
α_1: Constant	-65.9	-6.2	-86.5	-5.4	-10.9	-1.1	-19.6	-1.7	-38.5	-3.4
α_2: Constant	1.5	8.0	-0.4	-3.8	-1.6	-16.6	-3.9	-16.6	-1.7	-2.0
α_3: Constant	-0.1	-1.4	-0.1	-1.3	-0.3	-7.8	-0.5	-11.5	-0.3	-4.0
α_4: Constant	-4.4	-7.0	-5.8	-6.0	0.4	0.5	-1.7	-2.6	-2.5	-3.3
α_5: Constant	5.7	6.4	8.6	5.6	-1.1	-1.0	1.3	1.2	2.9	2.5
α_6: Constant	5.4	5.1	5.6	3.4	1.4	1.4	1.2	1.1	2.9	2.9
β_1: Constant	55.5	9.6	130.6	10.3	42.9	7.3	116.6	15.5	89.4	3.1
Wife's age	-2.8	-2.1	25.7	7.4	-2.0	-1.4	-2.7	-1.2	2.3	1.4
Husband's age	-2.8	-1.9	-17.6	-5.6	1.1	0.6	-3.5	-2.8	-4.7	-4.4
Youngest child 0-6	0.5	0.1	6.8	0.7	-34.3	-6.5	15.4	1.8	-0.7	-0.1
β_2: Constant	-8.9	-7.9	-0.6	-0.8	5.7	10.6	25.9	14.3	9.6	1.9
Wife's age	-0.1	-0.4	0.0	-0.1	0.4	1.1	0.1	0.3	0.1	0.6
Wife's age^2	0.0	0.4	-0.2	-3.5	0.0	-1.0	0.0	-1.5	-0.1	-2.6
Wife's education	-0.3	-5.1	-0.8	-5.8	-0.2	-2.5	-0.8	-11.6	-0.6	-8.3
Southern Italy	-0.3	-5.7	-1.1	-7.4	-0.2	-2.0	0.1	2.2	-0.2	-3.0
Youngest child 0-6	0.0	-0.2	-0.9	-2.1	1.9	7.3	-0.7	-2.2	0.0	0.0
Numb. of children	0.4	1.9	-2.4	-11.8	0.3	2.7	-0.4	-2.7	-0.4	-2.7
β_3: Constant	-2.8	-7.8	-4.3	-6.4	-0.6	-1.7	-1.6	-3.8	-2.1	-5.4
Husband's age	-1.2	-4.5	3.9	5.9	0.0	0.0	0.5	1.2	0.6	1.7
Husband's age^2	0.2	5.3	-0.6	-6.9	0.0	-1.2	-0.1	-1.7	-0.1	-2.0
Husband's educ.	-0.2	-2.7	-0.6	-4.9	0.1	0.9	-0.6	-5.7	-0.4	-5.2
Southern Italy	0.0	-0.8	0.1	0.9	-0.2	-2.8	-0.1	-1.4	-0.1	-1.5
Youngest child 0-6	0.0	0.2	-1.3	-3.1	1.5	5.4	-0.7	-1.8	-0.1	-0.6
θ_1: 1(hours husband=0)	-6.4	-7.8	-5.7	-3.9	-1.8	-2.8	-0.8	-0.9	-3.0	-2.8
θ_2: 1(hours wife=0)	-5.1	-3.8	7.6	7.3	8.0	15.9	56.4	16.9	24.3	2.9
Contributions to class membership (base = class 1):										
Constant	-		0.2	3.23	0.45	7.53	0.99	17.9		
Factor 1	-		0.6	10.4	0.88	15.4	1.08	20.5		
Factor 2	-		0.07	1.29	0.05	1.03	0.06	1.22		
Factor 3	-		0.21	3.71	0.16	3.01	0.12	2.5		
Factor 4	-		0.7	11.9	1.01	17.4	0.74	14.4		
Class probability (average)	0.21	3.41	0.17	1.90	0.23	7.73	0.39	4.91		
Log-likelihood:	-7691.49									

Note: model estimated via EM algorithm. Convergence achieved after 150 iteration. Standard errors computed using 50 bootstrapped samples

Table 3 Labor supply elasticities for married couples

Women labor supply elasticties:	MNL	RCML	HSML	LCML
All women	.62	.64	.66	.89
	(.56 .67)			
Women from southern Italy	.78	.82	.84	1.16
	(.70 .85)			
Women with high education	.53	.55	.57	.76
	(.48 .59)			
Women without children	.65	.70	.71	.99
	(.59 .72)			
Women with 1 child (<6)	.55	.56	.57	.75
	(.47 .63)			
Women with 1 child (<15)	.60	.62	.64	.85
	(.54 .66)			
Women with 2 children (<15)	.58	.60	.61	.78
	(0.51 .64)			
Women with 3 children (<15)	.52	.54	.56	.72
	(.44 .60)			
Women cross elasticities	-.04	-.07	-.09	-.15
	(-.09 .02)			
Men labor supply elasticties:	**MNL**	**RCML**	**HSML**	**LCML**
All men	.16	.17	.18	.28
	(.14 .18)			
Men from southern Italy	.27	.25	.28	.46
	(.23 .31)			
Men with high education	.10	.11	.12	.19
	(.08 .13)			
Men without children	.23	.23	.26	.34
	(.20 .27)			
Men with 1 child (<6)	.13	.12	.12	.27
	(.10 .16)			
Men with only 1 child (<15)	.12	.13	.14	.24
	(.11 .14)			
Men with 2 children (<15)	.09	.10	.12	.23
	(.07 .12)			
Men with 3 children (<15)	.05	.06	.07	.13
	(.03 .07)			
Men cross elasticities	.04	.06	.02	.10
	(.01 .07)			

Note: Bootstrapped 95% confidence interval in parenthesis (1000 replications, percentile method).

Table 4 Labor supply reaction to the WTC

	Pre-reform	Post-reform			
		LCML	MNL	RCML	HSML
Women:					
0 hours	50.85%	48.32%	49.80%	49.81%	49.69%
Part-time	19.37%	20.22%	19.68%	19.75%	19.75%
Full-time	29.78%	31.46%	30.52%	30.44%	30.56%
Tot	100%	100%	100%	100%	100%
Men:					
0 hours	8.38%	9.12%	8.85%	8.88%	8.87%
Full-time	91.62%	90.88%	91.15%	91.12%	91.13%
Tot.	100%	100%	100%	100%	100%

Note: Our computation based on the selected sample from EU-SILC (2006)

Table 5 Gini index before and after the reform

	LCML	MNL	MLHS	RCML
Gini index before:	32.27%	32.27%	32.27%	32.27%
Gini index after:	31.06%	31.39%	31.47%	31.44%
$\triangle$	-1.21%	-0.88%	-0.80%	-0.83%

Note: own computations based on EU-SILC 2006. For the computation of the Gini index after the reform we used the "pseudo-distribution" approach as in Creedy et al. (2006).

Figure 1: variation in women participation rates for decile of gross household income

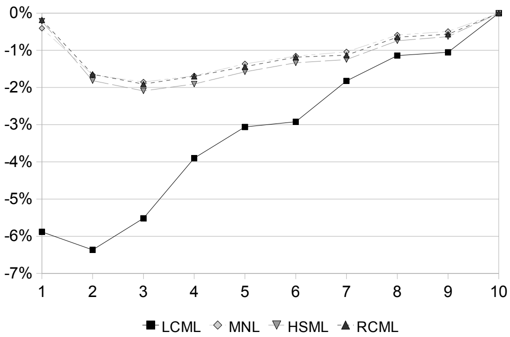


Figure 2: variation in women part time jobs for decile of gross household income

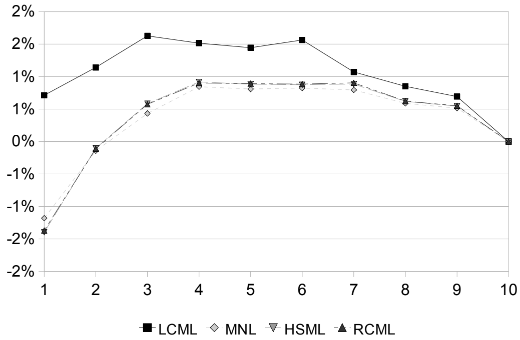


Figure 3: variation in women full time jobs for decile of gross household income

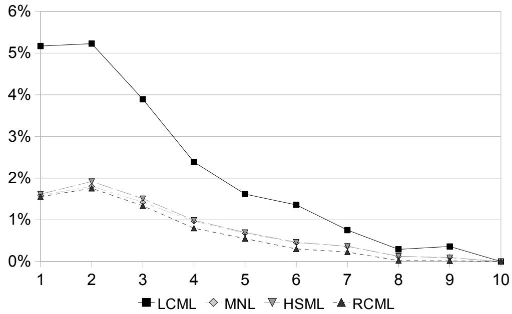
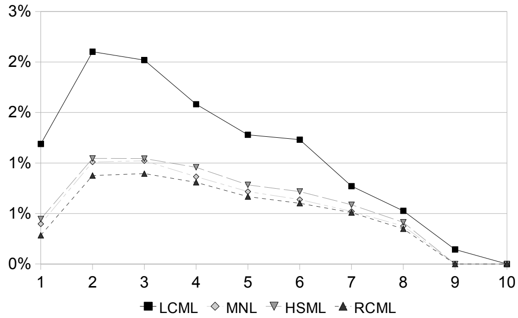


Figure 4: variation in men participation rates for decile of gross household income



References

Baldini, M., and Ciani E., (2009). Gli effetti distributivi delle principali riforme del sistema di tax-benefit italiano nel primo anno della XVI legislatura, CAPPaper, University of Modena and Reggio Emilia, 68.

Baldini, M., Pacifico, D., 2009. The recent reforms of the Italian personal income tax: distributive and efficiency effects. Rivista Italiana degli Economisti 1, 191-218.

Bargain, O., 2007. On modelling household labor supply with taxation. Working Paper No.07/11.

Bhat, C., 1997. An endogenous segmentation mode choice model with an application to intercity travel. Transportation Science 31 (1), 34-48.

Bhat, C., 2000. Flexible model structures for discrete choice analysis. Handbook of transport modelling 1, 71-90.

Blundell, R., Duncan, A., McCrae, J., Meghir, C., 2000. The labour market impact of the working families tax credit. Fiscal Studies 21 (1), 75-104.

Blundell, R., MaCurdy, T., 1999. Labor supply: A review of alternative approaches. Hand- books in Economics 5 (3 PART A), 1559-1696.

Chiappori, P., Ekeland, I., 2006. The micro economics of group behavior: general charac- terization. Journal of Economic Theory 130 (1), 1-26.

Colombino, U., Locatelli, M., 2008. Parameters Heterogeneity in a Model of Labour Supply: Exploring the Performance of Mixed Logit. ChilD working paper (21).

Creedy, J., Kalb, G., 2005. Discrete hours labour supply modelling: specification, estimation and simulation. Journal of Economic Surveys 19 (5), 697.

Creedy, J., Kalb, G., Scutella, R., 2006. Income distribution in discrete hours behavioural microsimulation models: An illustration. Journal of Economic Inequality 4 (1), 57-76.

Dempster, A., Laird, N., Rubin, D., et al., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B (Methodological) 39 (1), 1-38.

Greene, W., 2001. Fixed and Random Effects in Nonlinear Models. Working Papers.

Greene, W., Hensher, D., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. Transportation Research Part B 37 (8), 681-698.

Haan, P., 2006. Much ado about nothing: Conditional logit vs. random coefficient models for estimating labour supply elasticities. Applied Economics Letters 13 (4), 251-256.

Haan, P., Uhlendorff, A., 2007. Intertemporal Labor Supply and Involuntary Unemployment. Institute for the Study of Labor (IZA).

Heckman, J., 1974. Shadow prices, market wages, and labour supply. Econometrica 42 (4), 679-694.

Heckman, J., Singer, B., 1984. A method for minimizing the impact of distributional as- sumptions in econometric models for duration data. Econometrica 52 (2), 271-320.

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

McFadden, D., 1973. Conditional logit analysis of qualitative choice models. Frontiers of Econometrics, ed. P. Zarembka. New York: Academic Press.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete responses. Journal of Ap- plied Econometrics 15 (5), 447-470.

Pacifico, D., 2009. Estimation of a latent class discrete choice panel data model via Maximum Likelihood and EM algorithms in Stata. CAPPaper, University of Modena and Reggio Emilia 80.

Roeder, K., Lynch, K., Nagin, D., 1999. Modeling Uncertainty in Latent Class Membership: A Case Study in Criminology. Journal of the American Statistical Association 94 (447), 766-767.

Swait, J., 1994. A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data. Journal of Retailing and Consumer Services 1 (2), 77-89.

Thompson, B., Daniel, L., 1996. Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. Educational and Psychological Measurement 56 (2), 197-208.

Train, K., 2003. Discrete choice methods with simulation. Cambridge Books.

Train, K., 2008. EM Algorithms for Nonparametric Estimation of Mixing Distributions. Journal of Choice Modelling 1 (1).

Van Soest, A., 1995. Structural models of family labor supply: a discrete choice approach. The Journal of Human Resources 30 (1), 63-88.

Vermeulen, F., Bargain, O., Beblo, M., Beninger, D., Blundell, R., Carrasco, R., Chiuri, M., Laisney, F., Lechene, V., Moreau, N., et al., 2006. Collective models of labor supply with nonconvex budget sets and nonparticipation: A calibration approach. Review of Economics of the Household 4 (2), 113-127.

Wrohlich, K., 2005. Labor Supply and Child Care Choices in a Rationed Child Care Market. Institute for the Study of Labor (IZA).

Wu, C., 1983. On the convergence properties of the EM algorithm. The Annals of Statistics 11 (1), 95-103.