Modelling Unobserved Heterogeneity in Discrete Choice Models of Labour Supply

Pacifico, Daniele

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

15 November 2009

Modelling unobserved heterogeneity in discrete choice models of labour supply

Daniele Pacifico

(Preliminary, please do not quote!)

Abstract

The aim of this paper is to analyse the role of unobserved heterogeneity in structural discrete choice models of labour supply for the evaluation of tax-reforms. 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. Given the relative big set of parameters that enter in labour supply models, many researchers have to reduce the role of unobserved heterogeneity by specifying only a small set of random coefficients. However, this simplification affects the estimated labour supply elasticities, which then might hardly change when unobserved heterogeneity is considered in the model. In this paper, we present a new estimation method based on an EM algorithm that allows us to fully consider the effect of unobserved heterogeneity nonparametrically. Results show that labour supply elasticities do change significantly 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 labour supply reactions and post-reform income distribution do change significantly when unobserved heterogeneity is fully considered.

Jel classification: J22, H31, H24, C25, C14
key words: Labour supply, discrete choice model, latent class models, EM algorithm, mixed logit, random coefficients, working tax credit.

Introduction

Structural discrete choice models of labour supply are a useful tool for the *ex-ante* evaluation of labour 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 to handle highly non-convex budget sets and the non-participation choice easily[1]. As Blundell and MaCurdy (1999) point out, the discrete approach has to be preferred to other – continuous – labour supply specifications because of its overall flexibility, in particular when the aim is the *ex-ante* evaluation of a specific tax-reform. Modelling labour supply responses using a discrete approach has become increasingly popular in recent years. Earlier works that explore this method are those from Van Soest (1995), Keane and Moffitt (1998) and Blundell et al. (2000). The idea behind these earlier papers, which has now become standard in the literature, is to simulate real consumption over a finite set of alternative of leisure given the actual tax-benefit system. Under the hypothesis that agents choose the combination of leisure-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. Hence, what are estimated within this framework are the parameters of the direct utility function and not typical labour supply Marshallian functions, as in other (continuous) approaches. As for the rule of unobserved heterogeneity in discrete choice models of labour supply, 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 estimates nonparametrically using a latent class approach *a la* Heckman and Singer (1984). The idea is to assume a discrete distribution for the unobserved heterogeneity and to estimate the mass points and the population shares along with the other parameters of the utility function. Recent examples are from Haan (2006), Haan and Uhlendorff (2007), Wrohlich (2005), Bargain (2007) and Vermeulen et al. (2006). However, no matter the approach used, unobserved heterogeneity has always been assumed to affect only a relative small set of parameters, in particular those that mainly define the marginal utility of consumption and/or the marginal utility of leisure. The reason of this simplification does not lay on a specific economic theory but on the computational problems that normally arises with standard gradient-

[1] Within a discrete choice framework, the direct utility function already includes the budget constraint so that the optimisation problem does not need to be solved empirically. Hence, for the same reason, also the non-participation choice – which normally has to be treated separately, being a corner solution of the optimisation problem – can easily enter in the analysis.

base maximisation algorithms like Newton-Raphson or BHHH. Indeed, labour supply models contain a relatively high set of parameters so as to better explain how labour supply behaviour is related to the tax system. Moreover, the presence of random coefficients significantly changes the shape of the likelihood function increasing the probability of many local maxima. Hence, it follows that the higher is the number of parameters specified as random, the more difficult is the numerical computation of the gradient, which implies, in turns, a more instable Hessian with the related probability of singularity at some iterations. For this reason, the number of random parameters in labour supply models has always been kept small and this clearly curtails the role of unobserved heterogeneity. Hence, depending on the size of unobserved heterogeneity and on the number of coefficients specified as random, post-estimation results - as elasticities or other measures - might not differ significantly from those obtained without accounting for unobserved taste heterogeneity. Haan (2006) proved that no matter the way the researcher accounts for unobserved heterogeneity - parametrically or nonparametrically with just few random parameters - the subsequent labour supply elasticities do not change significantly with respect to the base model without unobserved heterogeneity. Haan's findings are actually confirmed by the evidence provided in this paper although we show 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 very different elasticities of labour supply. These finding is particularly important for the applied research whose aim is to evaluate empirically the labour supply reaction after tax-reforms. Indeed, different elasticities of labour supply imply different policy prescriptions 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 Expectation-Maximisation (EM) recursion that allow us to both speed-up estimation and ensure convergence. EM algorithms were introduced in the literature as a method to deal with missing data problems but they turned out to have an intuitive appeal for the estimation of latent class models where the class membership is the missing information. 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 travelling 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 lay on its overall stability. Moreover, as well explained in Train (2008), EM algorithms represent a relative easy solution for the nonparametric estimation of mixing distributions. The aim of this paper is hence twofold: firstly, we propose a new EM recursion for the nonparametric estimation of latent class discrete choice models that is quickly implementable, ensure convergence and speed-up estimation; secondly, we show that - in our data - unobserved heterogeneity affects post-estimation results only if a large set of parameters is assumed to be random. Our empirical analysis is based on the European panel on income and living conditions (EU-SILC). The empirical analysis is carried out in two main steps. Firstly, we estimate labour supply elasticities using different specifications of unobserved taste heterogeneity and show that they can differ significantly depending on the way unobserved heterogeneity is specified. Then, 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 labour supply elasticities can lead to different results in terms of labour supply reaction to tax reforms, different welfare changes and different post-reform income distributions. This paper is structured as follows. In section 1 we present the basic discrete choice model of labour supply. Section 2 shows how unobserved heterogeneity has been considered in this literature. Section 3 presents an overview of the EM recursion. Section 4 comments on the estimated utility parameters and compare elasticities across the various specifications of our model. Section 5 contains the simulation and the evaluation of the introduction of a UK-stile working tax credit schedule for Italy. Section 6 concludes.

The basic econometric model without unobserved heterogeneity

In this section we develop the econometric framework for the basic structural labour 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[2]. We assume that each household has a limited set of work alternatives and that spouses choose simultaneously the combination that maximise a joint utility function, which is defined over the household disposable income and the hours of work of either spouses[3]. 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

[2]Collective models of labour supply are much more appealing but the literature has not developed a well-accepted framework yet. In particular, the collective model has to be simplified in other directions and disputable assumptions are needed for the identification of the sharing rule parameter. See Chiappori and Ekeland (2006).

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

of the stochastic component is made. This is the base for the computation of the likelihood function. To be formal, let $\boldsymbol{H_j} = [hf_j\,; hm_j]$ be a vector of worked hours for alternative *j*, *hf* for female and *hm* for male. Let $y_{i,j}$ 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:

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

Where $\xi_{i,j}$ 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:

$$y_{i,j} = w_{i,f}hf_j + w_{i,m}hm_j + nly_i + TB(w_{i,f};\, w_{i,m};\, \boldsymbol{H}_j;\, nly_i;\, \boldsymbol{X_i}) \tag{2}$$

Where $w_{i,f}$ and $w_{i,m}$ are the hourly gross wages from employment for women and men respectively; nly_i is the household non-labour income and the function *TB(.)* represents the tax-benefit system, which depends on the gross wage rates, hours of work, household non-labour income and individual characteristics. It is worth to notice 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[4]. Following Keane and Moffitt (1998) and Blundell et al. (1999), the utility above is defined as a second order polynomial with interactions between the wife and the husband terms:

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

In order to introduce individual characteristics in the utility, 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 $\boldsymbol{H_j} =$

[4]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.

$[hf_j\,; hm_j]$ is given by[5]:

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

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

$$
LL = \sum_{i=1}^{N} log \prod_{k=1}^{K} (Pr(\boldsymbol{H} = \boldsymbol{H_j} \mid \boldsymbol{X_i}))^{d_{i,k}} \tag{6}
$$

Where $d_{i,k}$ is a dummy variable equals to one for the observed choice and zero otherwise. Importantly, it has been shown that the rounding error created by the categorisation of the worked hours does not create identification problems even if the true model is defined in the continuous time[6]. 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 underlay the standard conditional logit specification. The first one is about the stochastic components that enter the utility of each alternative, which are assumed to be independent across alternatives. The second assumption is that unobserved individual characteristics do not affect the response to variations in observed attributes. Finally, there is the assumption of error variance-covariance homogeneity which implies that the extent of substitutability among alternatives is the same across individuals. One prominent effect of these assumptions is the well-know property of *independence from irrelevant alternatives* (IIA) at the individual level. This property can be very restrictive in our labour supply framework. 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. The next section introduces different models that have been used in the labour supply literature in order to reduce the extent of the IIA property by relaxing one or more of the assumptions listed above.

Modelling unobserved heterogeneity in preferences

The literature has developed several models that relax the IIA property of the multinomial conditional logit. The random coefficient mixed logit is prob-

[5]See McFadden (1973)

[6]See Flood and Islam (2005).

ably the most important among numerous innovations because of its overall flexibility[7]. The idea that underlines this specification is that agents have different unobserved tastes which affect the individual response to given attributes. In other words, the parameters that enter the utility are not fixed across the population - like in the traditional multinomial logit model - but vary randomly with a given, unknown, distribution. In empirical works, the analysts specify a parametric distribution for this unobserved taste variability and its moments - normally the means and the standard deviations - are estimated along with the other preference parameters. Clearly, there is a great freedom in the choice of different densities and different alternatives can be tested. Common choices are the normal density, the log-normal or the triangular one. However, any parametric specification has several drawbacks implied in 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 mixed logit models is simply practical. Indeed, since the analyst does not observe the individual's tastes completely, the (conditional) probability of making the observed choice has to be integrated over all possible value 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 hard to compute, even with simulation methods. For this reason, the choice of many researchers is to reduce the number of random parameters so as to keep the estimation feasible. To be formal, it is convenient to rewrite the direct utility function of equation 3 in matrix form. In particular, let the utility of choice j for agent i be:

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

With $\boldsymbol{W}_{i,j} = (y^2_{i,j},\, hf^2_j,\, hm^2_j,\, hfhm_j,\, y_{i,j}hf_j,\, y_{i,j}hm_j)'$;$\boldsymbol{G}_{i,j} = (y_{i,j},\, hf_j,\, hm_j)'$ and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being the subsequent vectors of coefficients as in equation 3. Assume now the set of parameters in vector $\boldsymbol{\beta}$ to be random:

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Theta}\boldsymbol{X}_i + \boldsymbol{\Omega}\boldsymbol{\vartheta}_i \qquad E(\boldsymbol{\vartheta}_i) = \mathbf{0},\ Var(\boldsymbol{\vartheta}_i) = \boldsymbol{\Sigma} = diag(\sigma_1, \sigma_2, \sigma_3) \tag{8}$$

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, $\boldsymbol{\vartheta_i}$ the unobserved individual taste, $\boldsymbol{\Omega}$ a lower triangular matrix to be estimated

[7] See McFadden and Train (2000).

and $\boldsymbol{\Omega\Sigma\Omega'}$ defined as the variance-covariance matrix of $\boldsymbol{\beta}_i$. 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:

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

Where $Pr(H_i = H_{i,j} \,|\, \boldsymbol{X_i}, \boldsymbol{\vartheta_i})$ is the *conditional* probability of choice j. 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:

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

Where the integrals are approximated by the empirical expectation over the R draws from the selected (multivariate) distribution of the unobserved tastes. A recent literature has suggested latent class logit models as a variant of the standard multinomial logit that resembles the mixed logit model described above. Latent class models accounts for unobserved heterogeneity nonparametrically and have been proposed so as to be not constrained with distributional assumptions as in the random coefficient mixed logit model. These nonparametric models have been 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 need to model the probability of belonging to each class along with the probability of the observed choice in each class. Let us assume that there are C latent classes in the population of interest. Following the recent labour supply literature, we assume that only the preference parameters in vector $\boldsymbol{\beta}$ of equation 6 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 probability that household i belonging to class c chooses alternative j is:

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

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 probabilities in their right range and to ensure that they sum up to one for every household[8]:

$$Pr(class_i = c \,|\, \boldsymbol{\Delta_i}) = \frac{exp(\boldsymbol{\Delta_i'}\boldsymbol{\gamma_c})}{\sum_{c=1}^{C} exp(\boldsymbol{\Delta_i'}\boldsymbol{\gamma_c})} \qquad c = 1, 2, ..., C \;;\; \boldsymbol{\gamma_C = 0} \tag{12}$$

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. as Roeder, Lynch, and Nagin (1999) point out, these characteristics, which sometimes are called "risk factors", have to be specified properly. However, in many applications, in particular those related with the labour supply literature, these "risk factors" normally collapse to just a simple scalar in order to simplify the analysis and to speed-up estimation. Finally, it is worth to notice that the C*th* parameter vector is normalised to zero to ensure identification. Given equations 11 and 12, the conditional probability that a (randomly) selected household i chooses alternative j is:

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

Hence, the likelihood for the whole sample is:

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

As Train (2008) points out, differently from standard mixed logit models, the primary difficulty with nonparametric models is computational (rather than conceptual). Indeed, standard gradient-based method for ML estimation becomes more and more difficult as the number of parameters rises. For labour supply models this is even truer given the relative big set of parameters needed to model accurately the household behaviour. Hence, the choice of many labour supply analysts is to reduce enormously both the number of possible classes and the number of parameters assumed to be different in each class. Actually, the set of parameters traditionally assumed to be random is the same whether the analysis is carried out parametrically (with random coefficients mixed logit models) or nonparametrically (with latent class models). This way of modelling heterogeneity in labour supply models, with just a very small set of parameters

[8]See Greene (2001).

assumed to be random, could partially justify Haan's (2006) claim who didn't find significance differences in the labour supply elasticities obtained when unobserved heterogeneity is introduced either parametrically or nonparametrically. We actually confirm Haan's findings though we go a bit further and show that when a full latent class model is estimated the subsequent labour supply elasticities do change significantly. We are able to estimate a full latent class model of labour supply by means of a new estimation method that is not completely based on a standard gradient-based optimisation process. Indeed, we developed a new EM recursion that ensures convergences and speed-up the computation. The next paragraph contains an overview of this algorithm.

An EM recursion for discrete choice models of labour supply

The EM algorithm was initially introduced to deal with missing data problems[9] although turned out to be a very good way to estimate latent class models, where the missing data is the class membership. The recursion is known as "E-M" because it consists of mainly two steps, namely an "Expectation" and a "Maximisation". As well explained in Train (2008), the term being maximised 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 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 maximised by standard gradient-based methods as Newton Raphson or BHHH. However, the same log-likelihood can be maximised by repeatedly updating the following recursion:

$$\begin{aligned}\boldsymbol{\eta}^{s+1} &= argmax_{\eta} \sum_i \sum_c C_i(\boldsymbol{\eta}^s) ln \cdot w_c(\boldsymbol{\eta}^s) \prod_j \left[P(\boldsymbol{H_{ij}}|\boldsymbol{X_i},\boldsymbol{\pi_c})\right]^{d_{ij}} \\ &= argmax_{\eta} \sum_i \sum_c C_i(\boldsymbol{\eta}^s) ln(L_i \,|\, class_i = c)\end{aligned} \tag{15}$$

Where $\boldsymbol{\pi_c} = (\boldsymbol{\beta_c}\,\boldsymbol{;}\,\boldsymbol{\alpha_c})'$, $\eta = (\boldsymbol{\pi_c}\,;\,\boldsymbol{w_c},\, c = 1, 2, .., C)$, w_c is the density of the missing data in the population computed as in eq.12, L_i is the joint likelihood of both the observed choice and the missing data and $C(\boldsymbol{\eta}^s)$ is the probability that household i belongs to class c, conditional on the observed choice and the previous value of the parameters. This conditional (posterior) probability, $C(\boldsymbol{\eta}^s)$, is the key future of the EM recursion and can be computed by means of Bayes' theorem:

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

[9] See Dempster et al. (1977).

Now, given that:

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

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*[10],

2. Form the *individual-specific* conditional probablities of class membership using eq.16,

3. *For each class*, maximise the *expected* 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 likelihood in eq.14 and get a new set of w_c, $c = 1, ..., C$:

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

   Importantly, if the class shares w_c , $c = 1, ..., C$ do not depend on individual characteristics, these shares are update as follow:

$$w_c^{s+1} = \frac{\sum_i C_i(\boldsymbol{\eta}^s)}{\sum_i \sum_c C_i(\boldsymbol{\eta}^s)},\ c = 1, ..., C \tag{20}$$

   However, if the class shares depend on individual characteristics:

   - compute the new parameters that specify the impact of the risk factors as:

$$\boldsymbol{\gamma}^{s+1} = argmax_{\boldsymbol{\gamma}} \sum_{i=1}^{N} \sum_{c=1}^{C} C_i(\boldsymbol{\eta}^s) ln \frac{exp(\boldsymbol{\Delta_i'}\boldsymbol{\gamma_c})}{\sum_c exp(\boldsymbol{\Delta_i'}\boldsymbol{\gamma_c})},\ \boldsymbol{\gamma_C} = \boldsymbol{0} \tag{21}$$

   - update $w_{ic}(\boldsymbol{\eta}^s)$, $c = 1, ..., C$ as:

$$w_{ic}^{s+1} = \frac{exp(\boldsymbol{\Delta_i'}\boldsymbol{\hat{\gamma}_c^{s+1}})}{\sum_c exp(\boldsymbol{\Delta_i'}\boldsymbol{\hat{\gamma}_c^{s+1}})}, \qquad c = 1, 2, ..., C\ ;\ \boldsymbol{\gamma_C} = \boldsymbol{0} \tag{22}$$

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

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

It is worth to notice that in each maximization, the probability of class membership $C(class_i = c|\boldsymbol{\Delta_i}, \boldsymbol{\gamma})$ enters the likelihood without unknown parameters to be estimated and can be seen as an individual weight. Hence, equation 18 defines a standard conditional logit model with *weighed observations* that can be estimated easily with respect to the maximisation 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[11]. With this model in hand, it is possible to estimate a full latent class model of labour supply without being conditioned neither to the number of parameters assumed to be random nor to the number of classes. Moreover, the estimation time drops significantly with respect to the time spent by standard gradient-based algorithm used for the estimation of mixed logit models (both parametric or nonparametric)[12]

Empirical findings

For our empirical analysis we use the 2006 Italian wave of the European Union panel survey on Income and Living Conditions. We focus on the main category of tax-payer, i.e. households of employed, and allow for a flexible labour supply for both the spouses. Drawing from the previous literature, all couples are excluded in which either spouse is aged over than 65, self employed, student, civil servant or retired. These former households might have a different behaviour in the labour market that cannot be completely explained by the standard trade-off between leisure and consumption. Hence, they are assumed to have a fixed labour supply and are not considered in the following analysis. 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 categorised 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

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

[12] Both the continuous-random coefficient mixed logit models and the latent class model a la Heckman and Singer (1984) are very time consuming. With about 30 parameters and 4000 observations, our Stata routines take about 6 hours to get convergence in our Intel quad-core PC with 4Gbs of RAM (and STATA 10.1 MP); Our EM recursion take less than 1 hour to get convergence for a model with 4 latent classes and 127 parameters.

time and overwork), which implies 18 different combinations for each household[13]. 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)[14]. In what follows, we first consider the three models introduced in sections 1 and 2. In particular, the first model is estimated without accounting for unobserved heterogeneity and is then a typical multinomial conditional logit (MNL) as explained in section 1; the second model is far the most common in the applied labour supply literature and it is normally referred as the random coefficients mixed logit (RCML), which allows for unobserved heterogeneity using a parametric assumption for its distribution. In particular, following the model introduced in the section 2, we allow the 3 coefficients of the linear terms of the utility to be random with independent normal densities[15]. We then estimate the means and the standard deviations of these coefficients along with the other preference parameters using Simulated Maximum Likelihood[16]. 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 way of accounting for unobserved heterogeneity is getting widespread and it is commonly defined as a nonparametric estimation of mixed logit models *a la* Heckman-Singer (HSML). The model is estimated via Maximum Likelihood and for each random parameter we estimate its mass points and its population shares. As in any latent class analysis, one primary goal is the definition of the proper number of latent classes. This is still a controversial issue in the literature and hence we move along the main framework which defines the right number of classes as a function of the Bayesian Information Criteria (BIC)[17].The next table shows the estimated parameters for these three models, along with the maximised log-likelihood[18]:

[13]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.

[14]See Baldini and Ciani (2009)

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

[16]See Train (2003).

[17]See Greene and Hensher (2003) and Train (2008).

[18]For the HSML model only 2 classes are chosen since the maximum likelihood estimation with three latent classes did not achieve convergence.

Table 1: Estimated utility parameters (1)

		MNL		RCLM		HSLM	
		Coeff	z	Coeff	z	Coeff	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	1.56
	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:	Constant (class1)					59.55	13.45
β_1:	Constant (class2)					63.31	17.11
β_2:	Constant (class1)					-0.83	-3.13
β_2:	Constant (class2)					-0.80	-3.45
β_3:	Constant (class1)					-1.73	-6.75
β_3:	Constant (class2)					0.70	-2.61
	prob (class1)					0.78	-5.18
	Log-Likelihhod:	-8069		-8050		-8043	
	Observations:	4000		4000		4000	

Note: RCLM estimated by Simulated Maximum Likelihood with 500 Halton Draws; the σs are the estimated standard deviations for the 3 random coefficients 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". § denotes dummy variables and † means that the variable is measured in terms of deviation from its mean. Annual disposable household income divided by 1000; *hf* and *hm* are divided by 10; The square of the hours of work is divided by 1000 whilst the interaction terms are all divided by 100. 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.

As the table shows, most the coefficients have the expected sign over the three specifications[19]. 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 increase at a decreasing rate for all the households in the sample and this result holds over the three specification[20]. If we now observe the maximised 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. Unfortunately, the RCLM and the HSLM are not nested and a comparison of the coefficients would be miss-leading. However, using the Bayesian Information criteria, we could conclude that the latent class specification dominates the RCLM 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 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 generalise 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 parameters, they could be highly time consuming and could not guarantee convergences[21]. Hence, the LCLM is estimated throughout the EM recursion outlined in the previous section. As for the number of latent classes, we adopt the Bayesian Information Criteria and select four latent classes:

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

[20] 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.

[21] 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 12.

Table 2: 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

Another important issue is the right specification of the "risk factors" that enter the probability of belonging from a given class. In order to account for as much information as possible in the definition of these risk factors, we performed a principal-component factor analysis of the correlation matrix of a set of variables thought to be helpful for the explanation of class memberships. According to the Kaiser criterion, we retained the first four factors because the related eigenvalues are higher than one. The next table shows the (rotated) factor loadings obtained with the varimax rotation. At it 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 principal factors are related to the wife's and the husband's health conditions respectively whilst the last principal factor capture the socio-economic status.

Table 3: 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 cronic deseases†	-0.02	**0.8**	0.03	-0.05
Husband's cronic deseases†	-0.04	0.09	**0.77**	-0.09

According to Thompson and Daniel (1996), the households' risk factors that enter in our probability model are computed by using the scoring coefficients obtained through a standard regression model. The next table reports the coefficients for the LCML model with four latent classes along with their (weighted) average across the four classes[22]. As it can be seen, the maximised log-likelihood

[22]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.

is significantly higher with respect to the other models and also the fitting significantly increases[23]. Looking at the sign (and magnitude) of the average coefficients, we can see that, an average, the economic implications related with this model are in line with those from the other specifications. Importantly, using the estimated 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 by households living in the southern Italy, with young children and with relatively youth parents. Class 3, instead, is mainly composed by the same type of households but living in the northern Italy. Interestingly, these households have, on average, a higher education with respect to those household in class 1 and are more likely to have their own house. Class 4, instead, is mainly composed by relatively aged households, with far less young children and with relatively worst parents' health conditions. As for the analysis of preferences in each class, we computed the marginal (dis)utility of income (and work) in every class and evaluated the results using the probability of class membership. Interestingly, on average, households that are more likely to belong from class 1 and 3 have the lowest marginal utility of income, which could be partially explained by the relative young age of both parents. However, the households with a highest probability to belong from class 1 - mainly located in the southern Italy - have a higher marginal disutility of work if compared with the other classes. In general, the LCML model incorporates in the estimation more information than the other specifications so that many analyses could be made in order to better understand the source of unobserved heterogeneity. However, we defer this to other – more applied – studies.

[23]Table 8 in the appendix shows the predicted and actual frequencies for each alternative over our four specifications.

Table 4: 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 (mean)	0.21	3.41	0.17	1.90	0.23	7.73	0.39	4.91		
Log-likelihood:	-7691.49			Observations	4000					

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

We now turn to the main issue of this paper and compute the (average) elasticities across the various specifications of our labour supply models. Following Creedy and Kalb (2005), we computed such elasticities numerically. It is worth to notice 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 surely are a useful measure of the labour supply behaviour implied in our estimated model and can be used to check if the different specifications lead to different policy prescription[24]. More-

[24]Indeed, different elasticities across the various specifications would imply different labour

over, in order to better understand the relationship between the labour supply behaviour of each household member, we computed elasticities for each spouse. Labour supply elasticities are computed as follows. Firstly, gross hourly wages are increased by 1% for either spouses 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 labour 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 labour supply elasticities for either spouses is 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 labour supply elasticities, we adopt the same strategies of Hann (2006). In particular, we computed 95% bootstrapped confidence intervals for the MNL labour supply elasticities and checked whether these elasticities differ significantly from those obtained with other specifications. The next table shows the (average) own elasticities derived from 1% increase in the gross hourly wages of either spouses. As it can be seen, women's elasticities are higher than men's elasticities. Women cross elasticities are not significantly different from zero whilst men's cross elasticities are relatively higher and positive. If we now look at the elasticities by socio-demographic characteristics, we can see that elasticities are higher for those households in southern Italy (which is the poorest part of the country) and for people with low education. Children reduce labour 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 fell 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, we cannot reject the hypothesis of different elasticities. In particular,

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 distribution of income.

these elasticities are significantly higher with respect to the others, meaning that households have a significantly more elastic labour supply.

Table 5: Labour supply elasticities for married couples

Women l. supply elasticties:	MNL	RCML	HSML	LCML
All women	0.62	0.64	0.66	0.89
	(0.56 0.67)			
Women from southern italy	0.78	0.82	0.84	1.16
	(0.70 0.85)			
Women with high education	0.53	0.55	0.57	0.76
	(0.48 0.59)			
Women without children	0.65	0.70	0.71	0.99
	(0.59 0.72)			
Women with only one young child (<6)	0.55	0.56	0.57	0.75
	(0.47 0.63)			
Women with only one young child (<15)	0.60	0.62	0.64	0.85
	0.54 0.66)			
Women with two young children (<15)	0.58	0.60	0.61	0.78
	(0.51 0.64)			
Women with three young children (<15)	0.52	0.54	0.56	0.72
	(0.44 0.60)			
Women cross elasticities	-0.04	-0.07	-0.09	-0.15
	(-0.09 0.02)			
Men l.supply elasticties:	**MNL**	**RCML**	**HSML**	**LCML**
All men	0.16	0.17	0.18	0.28
	(0.14 0.18)			
Men from southern italy	0.27	0.25	0.28	0.46
	(0.23 0.31)			
Men with high education	0.10	0.11	0.12	0.19
	(0.08 0.13)			
Men without children	0.23	0.23	0.26	0.34
	(0.20 0.27)			
Men with only one young child (<6)	0.13	0.12	0.12	0.27
	(0.10 0.16)			
Men with only one young child (<15)	0.12	0.13	0.14	0.24
	(0.11 0.14)			
Men with two young children (<15)	0.09	0.10	0.12	0.23
	(0.07 0.12)			
Men with three young children (<15)	0.05	0.06	0.07	0.13
	(0.03 0.07)			
Men cross elasticities	0.04	0.06	0.02	0.10
	(0.01 0.07)			

Note: Boostrapped 95% confidence interval in parentesis (1000 replications, percentile method).

These findings are relevant in particular for the applied literature. Indeed, discrete choice labour supply model have been estimated only using the RCML or the HSML so far and the estimated coefficients are then used to predict the labour supply behaviour after tax reforms. However, we have shown that if unobserved heterogeneity is not considered properly, the resulting elasticities might be significantly different, which in turn implies different welfare (and political) evaluations related to tax reforms[25]. 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 labour supply reaction to the introduction of a UK-style working tax credit in the Italian tax-benefit system and show that income distribution and labour supply implications are significantly different depending on the approach used.

Simulating a WTC for Italy

The aim of working tax credits is to incentive the participation in the labour market for low income households. 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 with a series of individual characteristics as the number of young children, the age, the actual number of working hours and the 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[26]. In particular, our WTC is composed of 5 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

[25]Indeed, depending on the magnitude of labour supply elasticities, a given reform may produce different results in terms of welfare changes and income inequality.

[26]See www.direct.gov.uk and http://www.litrg.org.uk/help/lowincome/taxcredits/workingtaxcredit.cfm for more details

paid depends on the gross household income. In particular, according with 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 who have less than 25 years old and 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 old who work for at least 30 hours are eligible. For married/*de-facto* couples the benefit is primarily computed on an individual base and then the actual amount paid is the highest among the two spouses. The effect of WTCs has always been a controversial issue in the applied literature. Blundell et al (2000) and Brewer et al. (2006) found that the UK WTC has slightly reduced the participation rate of married women in the UK and increased the participation rate of both men in couples and lone mothers. However, other country-specific studies had different findings. 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 labour 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 get 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. The next table shows these aggregate frequencies before and after the reform for each specification of our model. As it can be seen, the sign of the labour supply reaction is the same over the 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 margine, the highest incentive for those women who would like to participate in the labour market is for full-time jobs, although there are also positive incentive for part-time options. If we now turn on the differences among the four models, it could be seen that the MNL, the RCML and the HSML share a very similar labour supply pattern after the reform. However, according with the elasticities computed in the previous section, the labour supply reaction produced by the LCML model is significantly stronger with respect the other specifications.

Table 6: Labour supply reaction to the WTC

	OBSERVED	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).

In order to better understand the differences among the four models, the next figures show, for each decile of gross household income, the absolute difference in the average frequencies of each labour supply category before and after the reform. As expected, mainly households in the lowest decile change their labour supply behaviour. However, the overall pattern of labour incentives is quite different if we consider the LCML model with respect to the other three specifications - the MLN, the RCML and the HSML - which again 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 focus on the same incentives using the LCML specification we observe firstly a significant different magnitude and, secondly, also a different structure of incentives across the various decile. In particular, 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 for the four models, although the LCML model produce, again, a stronger reaction, in particular for low-income households.

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

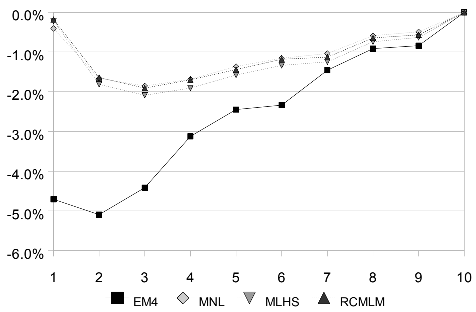


Figure 2: variation in women part time jobs for decile of gross household income

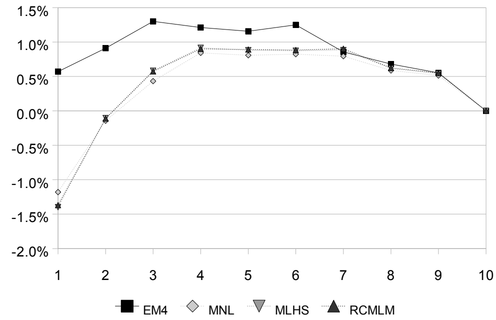


Figure 3: variation in women full time jobs for decile of gross household income

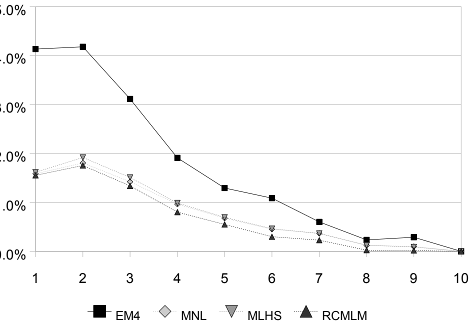
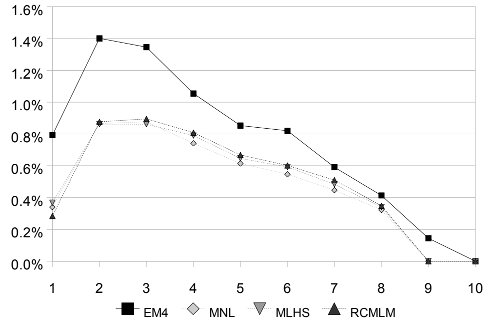


Figure 4: variation in men participation rates for decile of gross household income



In order to evaluate how the income distribution changes after the reform, we compute the Gini index befor and after the introduction of the WTC. As it can be seen in the next table, 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%). Moreover, for the other three specification, the reduction of the Gini index is similar and around -0.84%.

Table 7: Gini index before and after the reform

	LCLM	MNL	MLHS	RCMLM
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) .

Conclusions

The aim of this paper has been twofold. Firstly, we have shown that the way researchers account for unobserved heterogeneity can have an impact on the derived labour supply elasticities, which in turns implies that policy prescription related to particular tax-reform 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 implication, again, depends on the specification of unobserved heterogeneity. Secondly, we have provided an handful alternative to fully consider the effect of unobserved heterogeneity nonparametrically. In particular, we have proposed a easily-implementable EM algorithm that allow us to increase the number of random coefficients in the specification, ensure convergence and speed up the estimation process with respect to other gradient-based maximisation algorithms.

Appendix

Table 8: observed and predicted frequencis

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).

References

Baldini, M., and E. Ciani (2009): “Gli effetti distributivi delle principali riforme del sistema di tax-benefit italiano nel primo anno della XVI legislatura,” *CAPPaper*, 68.

Baldini, M., and D. Pacifico (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.

——— (2000): “Flexible model structures for discrete choice analysis,” *Handbook of transport modelling*, 1, 71–90.

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

Blundell, R., and T. MaCurdy (1999): “Labor supply: A review of alternative approaches,” *Handbooks in Economics*, 5(3 PART A), 1559–1696.

Brewer, M., A. Duncan, A. Shephard, and M. Suarez (2006): “Did working families' tax credit work? The impact of in-work support on labour supply in Great Britain,” *Labour Economics*, 13(6), 699–720.

Chiappori, P., and I. Ekeland (2006): “The micro economics of group behavior: general characterization,” *Journal of Economic Theory*, 130(1), 1–26.

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

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

Dempster, A., N. Laird, D. Rubin, 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.

Flood, L., and N. Islam (2005): “A Monte Carlo evaluation of discrete choice labour supply models,” *Applied Economics Letters*, 12(5), 263–266.

Greene, W. (2001): “Fixed and Random Effects in Nonlinear Models,” *Working Papers.*

Greene, W., and D. Hensher (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., and A. Uhlendorff (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., and B. Singer (1984): "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica*, 52(2), 271–320.

Keane, M., and R. Moffitt (1998): "A structural model of multiple welfare program participation and labor supply," *International Economic Review*, pp. 553–589.

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

McFadden, D., and K. Train (2000): "Mixed MNL models for discrete responses," *Journal of Applied Econometrics*, 15(5), 447–470.

Roeder, K., K. Lynch, and D. Nagin (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., and L. Daniel (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. (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., O. Bargain, M. Beblo, D. Beninger, R. Blundell, R. Carrasco, M. Chiuri, F. Laisney, V. Lechene, N. Moreau, 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.