Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA

ECONOMIA

CICLO XXII

SECS-P/06, SECS-P/03

TAX-BENEFIT MICROSIMULATION MODELS FOR THE EVALUATION OF PUBLIC POLICIES

Daniele Pacifico

Coordinatore Dottorato
Prof. Andrea Ichino

Relatore
Prof. Massimo Baldini

Esame Finale Anno 2010

Contents

Part I.

Preliminary Material

1. Acknowledgements

I would like to thank everyone who contributed to make possible the research contained in this thesis.

I am deeply grateful to my supervisor Massimo Baldini who always encouraged me and facilitated my research with precious tips and suggestions.

Andrea Ichino and Luca Lambertini, the coordinators of the Ph.D. program in Economics at University of Bologna also never denied their support. Richard Blundell, Costas Meghir Monica Costa Dias provided important feedback during my research periods at the University College London, UK.

Besides the contribution of my supervisor, the material in this thesis has benefited of the comments of many colleagues, senior academics and friends.

Comments on Chapter 3 were provided by several members of the IFS staff, in particular Monica Costa Dias and Mike Brewer and many colleagues as Miriam Hortas Rico and Alessia Russo.

Part of the material in Chapter 4 is related to an essay published in the journal "*Rivista Italiana degli Economisti*", which benefited of the comments of Paolo Bosi, Paolo Silvestri and an anonymous referee.

In writing Chapter 5, I benefited of the suggestions of Kenneth Train, Mike Brewer, Peter Haan and Monica Costa Dias. Part of the material included in the chapter was presented in several occasions, benefiting from several comments from the audience.

The support of the Ministero dell'Istruzione, dell'Università e della Ricerca, through the Ph.D. scholarship at University of Bologna is also gratefully acknowledged.

Finally, this thesis would not have been possible without the love of all my family - in particular of my mother, my father and my syster - and all my friends, in particular Maria Zindili, Francesco Zambelli and Sebastien Rippert.

2. Introduction

2.1. Static, dynamic and behavioural microsimulation models for the analysis of public policies

Microsimulation models are a useful tool for the ex-ante evaluation of specific tax-benefit reforms. They are called micro-simulation models since, differently from typical macroeconomic models, they preserve individual heterogeneity in the simulation exercise.

The literature has proposed several types of microsimulation models. The most common ones are known as static models. Their aim is to simulate in a very precise way the national (and local) system of taxes and benefits using large microdata surveys that are representative of the whole population. In particular, they recover gross earnings (if not observed) and then simulate taxes amounts and benefit entitlements for each unit in the sample in order to get the observed individual net disposable incomes. These kind of simulators are useful for the evaluation of the distributive impact of specific tax-benefit reforms. Indeed, once the tax-benefit system changes, the vector of net disposable income that is recovered by means of the simulation model will be different from the original one (which is computed according to the baseline tax-benefit system). Hence, the analyst can compute the income distribution before and after the change and identify winner and losers from the proposed reform.

These models are defined as "static" because they do not account for two important dimensions: individual dynamics (what happens in period t+1?) and behaviour (what is the individual reaction to a specific change in the economic environment?). According to these two dimensions, the literature has proposed dynamic microsimulation models and behavioural microsimulation models. Importantly, traditional dynamic microsimulation models do not account for behaviour in a structural way as a behavioural microsimulation model does. Indeed, they mainly work with transition probabilities (recovered from different source of data) and Markov processes so that a given event is not the result of an explicit optimal decision based on a structural microeconomic model.

As an example, assume a specific individual is observed working full time in period t. In period t+1, this individual will change his working time (say from full time to part-time) if, and only if, his/her specific transition probability - which is computed from a multivariate econometric model so that it depends on a set of observed individual characteristics - is higher then a random number drawn from a (typically uniform) probability distribution. The dynamic evolves in this way for many important events of the individual life, as fertility, marriage, child-bearing, education, health, disability, type of work, job effort, retirement, mortality, etc. Hence, the analyst can forecast the distribution of a specific characteristic - say net disposable income - over a long time span so that the effect of a specific reform today (as a pension reform) can be evaluated in the long run. The aim of behavioural microsimulation models, on the other hand, is to use the standard economic theory to predict specific changes in individual choices once a given element of the economic environment has changed.

As an example, assume the government introduces a new tax. Then, it is plausible that some people will change their behaviour in the labour market, say reducing the number of hours they work from a full time to a part time contract. According to this example, a behavioural microsimulation model uses a microeconomic model (typically with optimising behaviour subject to given budget constraints) to predict the overall change in the number of hours of work related to the change in the tax system. Of course, this behavioural change will take time to take place but the behavioural model do not consider the dynamic of this adjustment, it just gives the new final optimal choice. As these two examples have shown, behavioural and dynamic microsimulation models are intrinsically different and cannot be compared. However, thanks to the progress in computer technology, a new generation of models known as structural dynamic microsimulation models will be soon operative and will allow exploring the dynamics of behavioural reactions to specific changes in the economic environment over the life cycle[1].

The literature has already proposed a dynamic simulation model for the Italian case[2]. As for Italian behavioural microsimulation models, the work of Aaberge et al. (1998) and Aaberge et al. (2000) is particularly important, since these authors

[1] For a recent example of such models see van de Ven and Weale (2009).
[2] See Mazzaferro and Morciano (2008) for details.

have been some of the first in the international literature to propose a specific framework for the analysis of simulated behavioural responses to tax reforms using microsimulation models. However, the literature has proposed another approach for behavioural microsimulation models, which is based on the work of Van Soest (1995) and Keane and Moffitt (1998). As we will see, the two approaches are similar and their main difference rests on the way the individual choice set is constructed.

The main aim of this dissertation is to propose a behavioural microsimulation model that explores and expands the Van Soest's approach for the Italian case. Indeed, we propose a behavioural model that is able to account for many dimensions that are expected to influence the labour supply decision of many individuals. In particular, we allow for child-care expenditures, observed and unobserved individual preference heterogeneity for consumption and leisure, fixed costs of working and joint labour supply for married couples. The model is applied to the evaluation of the distributive and efficiency effects of the last three main reforms of the Italian personal income tax and important conclusions about the equity-efficiency trade off are discussed and analysed. Furthermore, we explore in deep detail the practical implications of unobserved preference heterogeneity for consumption and leisure in discrete models of labour supply, which is of interest to any kind of behavioural microsimulation model, no matter the approach used. In particular, we claim that the way researchers account for unobserved tastes in the econometric specification has an important impact on the subsequent estimates, and this - in return - implies that the efficiency and welfare analysis may change significantly depending how unobserved heterogeneity is introduced in the specification. Of course, this has important implications for the empirical literature, the aim of which is to evaluate specific tax-benefit reforms in order to derive policy recommendations.

In general, we believe that behavioural models are a very powerful tool for the evaluation of public policies and this thesis contributes to this literature exploring specific issues and proposing an innovative model for the Italian case. Obviously, much more can be done and several lines for future research are proposed in several occasions throughout the various chapters.

2.2. Behavioural microsimulation models and structure of this thesis

This thesis is dedicated to the analysis of behavioural microsimulation models for the evaluation of the effects of specific tax-benefit reforms on the labour supply decisions of the Italian population. As we have anticipated in the previous section, a behavioural microsimulation model contains a structural microeconomic model that explains the link between the two phenomena under analysis, that is the tax-benefit roles and the labour supply decision, in our specific case. Of course, the results of any simulation depend on the specific model used in the analysis and its assumptions. Hence, great care will be devoted to the specification of our model and to the analysis of its implicit and explicit assumptions.

Chapter 3 contains a detailed explanation of the model we propose, with a special focus on the econometric framework used for both the estimation and identification of the structural parameters. In particular, we will explore the performance of a labour supply estimation method based on a discrete choice set. The idea behind this approach is to work directly with preferences for leisure and consumption, instead of typical labour supply functions.

As it will be clarified, the main advantage of the discrete approach is the possibility to deal easily with non-convex budget sets, household labour supply and non-participation choices. Of course, these advantages let the discrete approach relatively suitable for policy evaluation purposes. We will use the papers from Blundell, Dancan, McCrae and Meghir (1999) and Brewer, Duncan, Shepard and Suarez (2006) as main references for the structural microeconometric model and several innovative elements will be taken into account with respect to previous Italian studies. In particular, we present a model that allows for errors in the predicted wages for non-workers, unobserved heterogeneity in preferences, unobserved monetary fixed costs of working and child-care demand.

The chapter concludes presenting and disussing the labour supply elasticities for married women and men that we computed throughout our model. Finally, an overview of the Stata® routine for the Simulated Maximum Likelihood estimation is presented in the appendix.

Chapter 4 contains an application of our behavioural model to the evaluation of the distributive and efficiency effects of the three most recent reforms of the Italian personal income tax. In order to show the functioning of our model, the analysis of these reforms will be carried out in three different stages. In the first one we will study the "pure" distributive effects of the reforms using the static part of our microsimulation model. In the second stage we will focus on the labour supply effects by means of the structural microeconometric model of household labour supply outlined in the previous chapter; finally, we will analyse the distributive effects of the reforms accounting for labour supply reactions.

Our findings confirm that the extension of the no-tax area had positive effects in terms of both redistribution and work incentives, and in the same time greater benefits for households with children improved income distribution but with negative effects on the labour supply of married women.

Finally, chapter 5 focuses on the role of unobserved preference heterogeneity in structural discrete choice models of labour supply. Within this framework, unobserved heterogeneity has been estimated either parametrically or nonparametrically through random coefficient models. However, several examples in the literature have shown that 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 tastes variability in empirical studies is constrained, since only a small set of coefficients is allowed to be random. However, this simplification may affect the estimated labour supply elasticities and the subsequent policy prescriptions.

Following this intuition, in this chapter we propose a new estimation method based on an EM algorithm that allows to fully consider the effect of unobserved heterogeneity nonparametrically. Results show that labour supply elasticities and policy prescriptions do change significantly only when the full set of coefficients is assumed to be random.

Moreover, we will 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 the post-reform income distribution can differ significantly depending on the specification of unobserved heterogeneity.

Part II.

Behavioural Microsimulation modelling

3. A behavioural microsimulation model for Italian married couples

3.1. Introduction

Discrete vs continuous structural labour supply models

Traditionally, structural labour supply models assume a choice set defined on any positive real number of worked hours. This is what Van Soest (1995) defines as the *continuous* approach. In this continuous framework, the agent chooses the best combination of consumption and leisure so as to maximise her utility function given a time and a budget constraint. Importantly, there are no constraints on the amount of leisure the agent can choose from: hours of leisure can be any real number up to the maximum amount available.

The literature has developed two different approaches for continuous labour supply. Often, a labour supply function is estimated relating hours worked with net-wage rates, non-labour incomes and individual characteristics. Then, indirect utility and expenditure functions are recovered by integration methods. Nevertheless, appropriate constraints on the parameters have to be imposed a priory so as to ensure duality conditions to hold. Moreover, in order to capture a relative wide range of labour supply behaviour, a reasonably flexible labour supply function is needed, with the subsequent difficulties during the integration procedure.

Another possibility in continuous microsimulation is to work directly with preferences with supply function derived from either a direct or an indirect utility function. Here the main problem is the tax schedule that enter the budget constraint, which can create several problems in the estimation stage.

In general, continuous microsimulation suffers of several problems no matter the approach followed. A first starting issue, for example, is how to recover the budget constraint for each possible level of labour supply. In continuous models, one or five minutes intervals of labour supply are needed for each individual, which means that with a standard total amount of time of 80 hours per week and thousands of individuals in the sample, this would be extremely time consuming[1].

Another complicated issue is the presence of a real tax-benefit system, which may give rise to highly non-linear and non-convex budget sets for most of the population of interest. This implies that feasible estimations require the linearisation of the budget constraint around the observed level of hours or the construction of search algorithms that compare the maximum utility on each linear segment of a

[1] Duncan and Stark (2000) have developed an algorithm in GAUSS that is able to recover budget sets more efficiently and accurately.

piecewise-linear budget constraint[2].

Moreover, considerable problems arises because of the simultaneity between net-wages and hours of work with the subsequent necessity to find appropriate instruments so as to ensure identification. Finally, other difficulties arise when the model try to allow for important extensions like unobserved preference heterogeneity, joint labour supply and non-participation.

As Creedy and Duncan (2002) point out, these criticisms make the continuous approach seldom used nowadays. Indeed, the recent literature has shown that a discrete approach to labour supply modelling solves most of the problems that are typically found in continuous models and allows for some important extensions.

The discrete model of labour supply is still based on the assumption of utility maximising agents as in the continuous approach but now the agent is constrained to choose from just few hour points instead of any possible hour in the real line. The utility is defined over income and leisure (hours of work) and any assumption is made *a priori* on the marginal (dis)utility of leisure (work) and income. If a stochastic component is added to the utility function, then the probability of a particular choice of hours of work can be derived and the likelihood function can be computed. In other words, what is estimated in the discrete approach are not the parameters of a classical Marshallian labour supply function but the parameters that define the shape of the utility function.

Given that the tax-benefit system enters the utility only indirectly through the consumption term, complicated tax schedules and non-convex budget sets do not represent a problem anymore. Moreover, any problems arise from the choice of the utility function given that the form of the probabilities depends on the assumptions on the utility stochastic component. Finally, the budget constraints have to be computed for just the few hour points the agent is constrained to choose from, and not for any possible level. Furthermore, the discrete approach allows for important extensions that are difficult to consider in the continuous model. Indeed, as it will be clarified later, wage unobserved components for non-workers, child-care demands, fixed costs of working, heterogeneity in preferences, non-participation and joint labour supply can be incorporated in the model in a very convenient way.

[2] A first generation of model linearized the budget constraint by computing the average net wage rate corresponding to the observed hours. Other subsequent models have elaborated algorithms that examine the full budget constraint when searching for the optimal level of labour supply, allowing for nonlinearities and nonconvexities. See Creedy and Duncan (2002) for references.

The main drawbacks of the discrete approach are the rounding errors produced when the choice set is discretised and the incomplete use of available information[3].

As Blundell and MaCurdy (1999) point out in their literature review of labour supply modelling, the discrete approach has to be preferred to other models because of its flexibility, in particular when the aim is the *ex-ante* evaluation of a specific policy reform. Modelling labour supply responses using a discrete approach has become increasingly popular in recent years, in particular when the aim of the analysis is the evaluation of a specific tax reform.

Earlier international works that explore this method are those from Van Soest (1995), Keane and Moffitt (1998) and Blundell et al. (2000). The econometric model used in these papers has now become standard in the literature and a similar version is also used in this work. Recent examples include: Brewer et al. (2006) who extend the paper of Blundell et al. (2000) to study the impact of the WFTC reforms in the UK, Breunig et al. 2008 who estimate the wage equation and the structural labour supply model simultaneously allowing for correlation between the random terms, Haan (2006), who studies the German case comparing the performance of a random coefficients specification with respect to the performance of a more simple model without unobserved heterogeneity, Labeaga et al. (2008) who study the impact of the Spanish tax reform on efficiency and social welfare. A very active centre that is specialised on microsimulation and labour supply in a discrete choice framework is at the Melbourne Institute of Applied Economics. I remand to Creedy and Kalb (2005) for a review of some of their papers.

For the Italian's case Aaberge et al. (2000) developed a model of labour supply allowing for different job types for each household; in their paper, job alternatives are defined over a continuum of wage rates, hours of work and other job characteristics. The analyst does not observe the opportunity set of each household so that the probability of choosing a particular job has to be weighted with the probability of receiving that particular job offer. Recently, Mancini (2007), developed a model that is closely related to the one discussed here in order to study the labour supply response to minimum income policies. Del Boca et al. (2005) study the impact of child-care rationing on female labour supply using a bivariate probit model for the joint decision of child-care and labour supply.

However, differently from the other models that have been developed for the

[3] The wider the hours categories used to discretise the choice set the bigger the rounding error, see Van Soest (1995) for this point.

Italian case, the model presented in this paper accounts for many innovative features simultaneously. Indeed, we estimate jointly the labour supply behaviour of married women and men allowing for errors in predicted wages, work-related monetary costs, child-care costs and unobserved random preferences for consumption and leisure.

Work-related monetary costs are important since they eliminate the reservation wage condition in estimation and, depending on how these costs are specified, they may help to relax the assumption of fixed wage rates[4].

Furthermore, we also account for endogenous child-care expenditures and unobserved heterogeneity in preferences. Both these features are relevant. From a practical point of view, child-care expenditures have a role similar to those of other fixed costs of working since they may help to eliminate the reservation wage condition for household with children during the estimation. Unobserved heterogeneity is also important so as to get unbiased estimates given the assumption made on the distribution of the utility function[5].

Finally, one more important difference from previous studies regards the way we take into account the endogeneity related to estimated wages for non-workers. Indeed, we integrate out wages by drawing randomly from their estimated distribution and weighting the likelihood when wages are not observed in the data.

This chapter is structured as follows. Section 2 presents the general framework for the structural microeconometric model. Section 3 presents the extensions to the basic model. Section 4 explains the data used for the empirical analysis. Section 5 describes the estimation procedure. Section 6 contains results from first stage regressions and section 7 discusses the estimates of the structural model. The appendix contains an overview of the Stata® algorithm coded for the estimation of the structural model.

[4]See Brewer et al. (2006) for this point.

[5]As it will be clarified later, the direct utility function is assumed to follow a type 1 extreme value distribution, which underlines the typical IIA (independence from irrelevant alternatives) property. The way unobserved heterogeneity is introduced can relax the extent of this assumption and this is important because such property could be particularly binding in our labour supply framework.

3.2. The structural model

In this section we develop the econometric framework for the empirical analysis. We focus only on married/*de facto* couples and do not consider singles[6]. In what follows we adopt a unitary framework to model the intra-household decision process, which means that the couple as a whole has to be considered as the decision maker. However, it is worth noting that a recent literature has shown that the unitary assumption is often rejected empirically. Nevertheless, a *collective* model of labour supply is far away to be a practicable model to be used for a detailed policy analysis. Moreover, a collective model has to be simplified in other directions and is based on discountable assumptions for the identification of the bargaining parameter[7].

For these reasons we follow the main literature and assume a unitary model of labour supply. This means that the two members of the couple simultaneously choose a particular combination of hours of work for both of them in order to maximise a joint utility function defined over the household net income and the hours of work of both the spouses.

As common in the literature, we assume that the gross wage rates are fixed and do not depend on the hours of work[8]. This implies that the hours of work uniquely define the household's gross income alternatives while the tax-benefit system uniquely defines the net household income alternatives. The decision is then taken given the tax-benefit system and the gross wage rates.

Under the assumption that the couple is utility maximising and that the utility is not deterministic, it is possible to recover the probability of a particular choice, which 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 married women and *hm* for married men. Let y_{ij} be the net household income 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:

$$U_{ij} = U(y_{ij},\, \boldsymbol{H_j},\, \boldsymbol{X_i}) + \xi_{ij} \tag{3.1}$$

[6]The labour supply model for single is based on the same assumptions and can be easily constructed and estimated.

[7]See Chiappori (1992) and Blundell et al. (2008) for a collective model of labour supply. Chiappori and Ekeland (2006) discuss on the identification assumptions in a collective framework.

[8]Some studies have found a part time pay penalty, see Manning and Petrongolo (2008). However there are several ways that relax this assumption that will be discussed in the next sections.

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. This component capture any couple-specific misunderstanding in the perception of the utility derived from a particular choice of hours and it can be seen as an optimisation error. The net-household income of household i when alternative j is chosen is defined as follows:

$$y_{ij} = \bar{w}_{if} hf_j + \bar{w}_{im} hm_j + nly_i + TB(\bar{w}_{if};\ \bar{w}_{im};\ \boldsymbol{H}_j;\ nly_i;\ \boldsymbol{X_i}) \tag{3.2}$$

Where $\bar{w}_{if}$ and $\bar{w}_{im}$ are the (fixed) hourly gross wages from employment for women and men respectively; nly_i is the household non-labour income and the function $TB(\bar{w}_{if};\ \bar{w}_{im};\ \boldsymbol{H}_j;\ nly_i;\ \boldsymbol{X_i})$ 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 noting that this function can be highly non-linear for most of the population of interest. Following Keane and Moffitt (1998) and Blundell et al. (2000), the observed part of the utility defined above is parametrised as a second order polynomial with interaction between the wife and the husband terms:

$$\begin{aligned} U(y_{ij}; \boldsymbol{H}_j; \boldsymbol{X_i}) \quad = \quad & \alpha_1 y_{ij}^2 + \alpha_2 hf_j^2 + \alpha_3 hm_j^2 + \\ & + \alpha_4 hf_j hm_j + \alpha_5 y_{ij} hf_j + \alpha_6 y_{ij} hm_j + \\ & + \beta_1 y_{ij} + \beta_2 hf_j + \beta_3 hm_j \end{aligned} \tag{3.3}$$

To introduce individual characteristics in the utility, the coefficients of the linear terms are defined as follows:

$$\beta_j = \sum_{i=1}^{K_j} \beta_{ij} x_{ij} + \nu_j \quad j \epsilon \{1, 2, 3\} \tag{3.4}$$

with the ν_j terms being the unobserved household preferences for both income and work, which are assumed to be independents and normally distributed.

The presence of these random terms is important for two reasons. On the one hand, they relax the IIA assumption which is implicit whenever the latent factor (here the utility gained from each alternative) follows a standard extreme value distribution[9]. On the other hand, they allow for heterogeneity in preferences in

[9]IIA is the acronym of Independence form Irrelevant Alternatives. This property is particularly restrictive in the labour supply framework. Consider a choice set initially defined by just two alternatives:

the model.

Under the assumption that the couple maximises her utility over a discrete set of alternatives and that the error term in the utility function follows a type-one extreme value distribution independent across alternatives, the probability of choosing a particular vector $\boldsymbol{H_j} = [hf_j\,; hm_j]$ is given by[10]:

$$\begin{aligned} Pr(\boldsymbol{H} = \boldsymbol{H_j} | \boldsymbol{X_i}) &= Pr[U_{ij} > U_{is}, \forall s \neq j] \\ &= \frac{exp(U(y_{ij},\, \boldsymbol{H_j},\, \boldsymbol{X_i}))}{\sum_{k=1}^{K} exp(U(y_{ik},\, \boldsymbol{H_k},\, \boldsymbol{X_i}))} \end{aligned} \tag{3.5}$$

Given the presence of unobserved components, it is necessary to integrate over their distributions to evaluate the likelihood function. For observation i the likelihood is:

$$\mathcal{L}_i = \int_{\boldsymbol{\nu}} \prod_{j=1}^{K} \left(\frac{exp(U(y_{ij}, \boldsymbol{H_j}, \boldsymbol{X_i}))}{\sum_{k=1}^{K} exp(U(y_{ik}, \boldsymbol{H_k}, \boldsymbol{X_i}))} \right)^{d_{ij}} \phi(\boldsymbol{\nu}) d(\boldsymbol{\nu}) \tag{3.6}$$

Where d_{ij} is a dummy variable equal to one for the observed choice and zero otherwise. Up to this point, we have not considered the problem of not observed wages for non-workers. The approach adopted here is to make an assumption on the wage generating process and to estimate the wage rate *before* estimating of the structural model of labour supply.

Of course, it would be more efficient taking into account the incidental truncation during the estimation of the structural model but the relative gain in efficiency is offset by an high increment in computational time[11]. We assume that wage for agent i is generated by the following selection process:

$$log(w_i) = \boldsymbol{X_{1i}}\beta + \epsilon_i \tag{3.7}$$

$$w_i \; is \; observed \Longleftrightarrow U_i^* = utility \; of \; work > 0$$

$$U_i^* = \boldsymbol{Z_i}\alpha + \varepsilon_i \tag{3.8}$$

$$\begin{pmatrix} \epsilon_i \\ \varepsilon_i \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{pmatrix} \sigma_\epsilon^2 & \rho \\ \rho & 1 \end{pmatrix} \right) \tag{3.9}$$

working full time and not working. The IIA assumption implies that introducing another alternative, say a part time job, does not change the relative odds between the two initial alternatives.

[10] See McFadden (1973) for a proof.

[11] See Keane and Moffitt (1998).

Where $\boldsymbol{Z}_i = [\boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}]'$ is a vector of individual characteristics. The assumptions on the wage generation process allows to estimate consistently the gross wage rate whenever it is not observed in the data. Then, the unobserved component of wages is integrated out from the likelihood during the labour supply estimation by drawing randomly from its distribution. This means that the likelihood changes as follows:

$$\mathcal{L}_i = \int_{\epsilon}\int_{\boldsymbol{\nu}} \prod_{j=1}^{K} \left(\frac{exp(U(y_{ij}, \boldsymbol{H_j}, \boldsymbol{X_i}))}{\sum_{k=1}^{K} exp(U(y_{ik}, \boldsymbol{H_k}, \boldsymbol{X_i}))} \right)^{d_{ij}} \phi(\boldsymbol{\nu})\phi(\epsilon)d(\boldsymbol{\nu})d(\epsilon) \tag{3.10}$$

Where the integration over ϵ takes place only when the wage is not observed. As anticipated in the introduction, the discrete approach allow considering important extensions to this basic model as fixed costs of working and child-care demand, which could be added in very convenient way to the basic model. This is the aim of the next section.

3.3. Extension to the basic model

The model outlined in the previous section is not able to replicate the data accurately. The main reason is that it does not take into account the characteristics of particular hour points that may help to eliminate the reservation wage condition[12].

There are several ways to account for these problems. For example, Aaberge et al. (2000) allow for different job offers for each individual with job alternatives defined over different combinations of wage rates and hours of work. This specification explicitly allow for different characteristics of each level of labour supply since it assume a distribution of job offers that puts more weights on particular hours categories. However, it increases the computational burden since the choice set becomes actually infinite, which then requires sampling methods so as to let the estimation feasible[13].

Another common approach is to allow for different characteristics of particular hours points by using *ad-hoc* penalties in the utility function[14]. This method may serve to account for different hours characteristics but it is not clear whether it could eliminate the reservation wage condition. Moreover, this procedure has the additional problem that the estimated coefficients of the penalty variables are measured in term of utility and do not represent monetary values.

The approach we follow is the one used in Brewer et al. (2006). The idea is that labour supply is often constrained implying a loss when the agent has to choose an alternative that is not exactly the one she would choose without constraints. Hence, different hour alternatives may imply different costs. These costs can be both psychological and physical but in both cases they can be quantified in monetary terms. Intuitively, these costs are on average higher for the choice between non-participation (zero hours of work) and participation but there could be also an additional cost for the choice between part-time alternatives and full-time alternatives.

Following this idea, we consider the characteristics of different discrete points

[12] Indeed, it has been found that the basic model systematically over-predict part-time alternatives and non-participation. See Van Soest (1995) for a discussion.

[13] The authors approximate the infinite choice set with a sample of weighted alternatives, with the weighting depending on the sampling scheme from the univariate densities of wages and hours. This make their model somewhat close to the multinomial logit used in this paper. See Aaberge et al. (2000) for details.

[14] This approach has been followed by several authors, see Mancini (2007), Haan (2006) and Van Soest (1995)

by estimating the monetary cost of working for three groups of discrete points. In particular, we allow for different work-related costs by distinguishing among non-participation, part time and full time alternatives. These work-related costs are modelled as fixed, unobserved costs directly subtracted from net income at positive working hours with an additional cost whether the agent chooses to work full time[15].

Differently from other papers that use *ad-hoc* penalties, the approach we follow allows the estimation of values that are indicative of the real monetary cost of choosing a given amount of worked hours. More importantly, this method may serve to relax the assumption that wage rates are fixed across alternatives, a point that actually drives also the specification of Aaberge et al. (2000). Finally, several studies have shown that netting out monetary costs of working from net income may also have the positive effect of leading to estimated preferences that are more likely to be convex[16].

Formally, fixed costs can be defined as follows[17]:

$$FC(hf_j, \boldsymbol{Z}) = \boldsymbol{Z}_1\boldsymbol{\gamma_1}\cdot 1\{hf_j > 0\} + \boldsymbol{Z}_2\boldsymbol{\gamma_2}\cdot 1\{hf_j > 30\} \tag{3.11}$$

Where $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are vectors of individuals characteristics, γ_1 and γ_2 are vectors of parameters that are estimated jointly with the other structural parameters and $1\{-\}$ are binary indicators that take value one whenever the argument inside the brackets is true.

To take into account child-care costs we adopt a different strategy. As pointed out in Del Boca et al. (2005), Italy has an objective lack of data on child-care usage and child-care costs. In order to overcome this problem, we recover information on child-care costs from another database. In particular, we computed the hourly price of child-care for different groups of households and for each group we approximated the distribution of the hourly price of child-care by a 4 point mass distribution whenever the household is observed buying formal child-care.

Given that households with working mother are more likely to buy formal child-care, we take into account a possible selection bias by computing the proportion of households that use formal child-care for both working and non-working mothers.

[15] Identification of these costs follows from the exclusion of the non-participation category.

[16] See Heim and Meyer (2004).

[17] As in Blundell et al. (2000) and Brewer et al. (2006), we assume positive costs of working only for female.

We do not consider any other possible source of selection bias, which implicitly means that households that are not observed buying formal child-care would pay exactly the same amount as households observed buying formal child-care. Finally, we estimate the statistical relationship between hours of work and hours of child-care for different groups of households defined according to the number of children and their age.

Whit this information on child-care costs and child-care usage it is possible to approximate the weekly cost of child-care for different alternatives of working hours in the original database. This cost is then subtracted from net income at any possible choice of hours and the price of child-care is then integrated out from the likelihood in order to account for unobserved quality.

To be formal, define a child-care cost function as:

$$CC(hf_j, p_c, \boldsymbol{X_i}) = E[h_{cc}|\boldsymbol{X_i}, hf_j]{\cdot}p_c \tag{3.12}$$

Where p_c is a particular price for an hour of child-care and $E[h_{cc}|\boldsymbol{X_i}, hf_j]$ is the expected hours of child-care for a particular household's group given the choice hf_j. Like work-related costs, child-care costs enter the model as a once off cost directly subtracted from net income at any possible choice of hours. Defining a total cost function as:

$$TC_i = CC(hf_j, p_c, \boldsymbol{X}_i) + FC(hf_j, \boldsymbol{Z}_i) \tag{3.13}$$

the utility function changes as follows:

$$U_{ij} = U(y_{ij} - TC_i,\ \boldsymbol{H_j},\ \boldsymbol{X_i}) + \xi_{ij} \tag{3.14}$$

and the likelihood for observation i becomes:

$$\mathcal{L}_i = \sum_{s=1}^{5} P(p_c^s|\boldsymbol{X_i}) \int_{\boldsymbol{u}} \prod_{j=1}^{K} \left(\frac{exp(U(y_{ij} - TC_{ij}, \boldsymbol{H_j}, \boldsymbol{X_i}))}{\sum_{k=1}^{K} exp(U(y_{ik} - TC_{ik}, \boldsymbol{H_k}, \boldsymbol{X_i}))} \right)^{d_{ij}} \phi(\boldsymbol{u})d(\boldsymbol{u}) \tag{3.15}$$

Where $\boldsymbol{u} = (\epsilon, \nu_1, \nu_2, \nu_3)$ collects all the random terms, $P(p_c^s|\boldsymbol{X_i})$ is the probability that household i faces a price of child-care p_c and d_j is a dummy that picks up the observed choice.

The likelihood above is difficult to estimate since it requires the computation of a four dimensional integral. Following Train (2003), we apply simulation methods to approximate these integrals. In particular, we use Halton sequences instead of traditional random draws from the densities since they ensure a more complete coverage of the integration support, which implies that a smaller number of draws is needed to reach consistent estimates. The simulated log-likelihood is:

$$\mathcal{L}_i = \sum_{s=1}^{5} P(p_c^s|\boldsymbol{X_i}) \frac{1}{R} \sum_{r=1}^{R} \prod_{j=1}^{K} \left(\frac{exp(U(y_{ij} - TC_{ij}^s, \boldsymbol{H_j}, \boldsymbol{X_i}, \boldsymbol{\nu_r}))}{\sum_{k=1}^{K} exp(U(y_{ik} - TC_{ik}^s, \boldsymbol{H_k}, \boldsymbol{X_i}, \boldsymbol{\nu_r}))} \right)^{d_{ij}} \tag{3.16}$$

Where R is the number of Halton draws used. The routine we coded to maximise this likelihood is discussed in the appendix.

3.4. Data

The main source of data is the Survey of Household Income and Wealth (SHIW) conducted by the Bank of Italy every two years. The survey has both a cross section and a panel dimension. It collects very detailed information on earnings as well as social and demographic characteristics. This survey has been widely used for labour supply analysis and policy evaluation[18].

In the present study we use the cross sectional survey for the year 2002. The database is representative of the whole Italian population and contains about 21,000 observations and 8,000 households. Since the model presented in the previous section is not appropriate to describe the labour supply decisions of any kind of household, we focus only on a selected sub-sample of the whole population. In particular, as standard in the literature on labour supply, we do not consider couples with either spouses are aged over 60 years, self-employed, involved in a full time education programs or serving the Army. Couples with self-employed spouses are omitted because it is difficult to estimate their budget constraint correctly. The other excluded couples are omitted because they might have a behaviour in the labour market that is not characterised by just the traditional trade-off between leisure and income.

Very detailed information about net income and wealth is provided in the survey. To recover gross incomes we use a modified version of the arithmetic tax-benefit microsimulation model called *MAPP02*. This model has been developed at the University of Modena and Reggio Emilia[19] and is able to generate gross incomes, benefit entitlements and tax amounts for each household in the data. *MAPP02* has been adapted to make it suitable for the present study. In particular, the Stata® modules of *MAPP02* have been modified to generate different vectors of taxes (positives and negatives) and net individual incomes for any possible combination of worked hours among which the couple can choose from.

As explained in the previous section, we also use another database to recover information about child-care costs and child-care usage. In this case, the source of data is the survey "MULTISCOPO" 1998 on Households and Childhood Conditions, which is conducted by the Italian national institute of statistics (ISTAT). This survey is relatively old but it is the only one that contains detailed information

[18] See Del Boca et al. (2005), Aaberge et al. (2000), Mancini (2007), Brandolini (1999), Baldini et al. (2002), Baldini and Pacifico (2008).

[19] See Baldini (2004).

on child-care expenditure, hours of child-care and hours of work. Unfortunately, the information on child-care expenditures is registered only for children with less than 7 years so that we are able to compute child-care costs only for those couples who have young children.

Nevertheless, this is not a great limitation since the government provides universal education for older children. The two databases are relatively similar and both are representative of the same population.

3.5. Estimation procedure

In this section we comment on the procedure used for the estimation of the structural labour supply model. The estimation process is divided in steps, following the natural development of the model outlined in the previous sections.

The first step is the definition of the relevant sub-sample and the re-arrangement of the information contained in the two databases so as to gather all the information needed for the analysis. This implies the alignment of the two sources of data so that the information can be compared.

The next step is to use the tax-benefit simulator so as to recover gross wages for those observations observed working. The information on the number of worked months and the average hours of work per week is used to recover the gross hourly wages. Gross hourly wages for unemployed workers are estimated making use of the wage generating process outlined above, which leads to a classical Heckman selection model. This is done separately for both the spouses in the couple. Using post-estimate results from the wage equations it is possible to predict hourly gross wages for non-workers. Following the wage model outlined above, predicted wages for non-workers could be computed as:

$$E[ln(w_i)|ln(w_i) \textit{ is not observed}; \boldsymbol{X_i}] = \boldsymbol{X}_i\hat{\beta} - \hat{\sigma}_\epsilon\hat{\rho}\lambda(\boldsymbol{Z}_i\hat{\alpha}) \tag{3.17}$$

Where $\lambda(\boldsymbol{Z}_i\alpha) = \frac{\phi(\boldsymbol{Z}_i\alpha)}{\Phi(\boldsymbol{Z}_i\alpha)}$ is the *mills ratio* computed according to equation [7]-[9]. However, using just the predicted wages for non-workers would lead to inconsistent estimates as long as wages are endogenous and predicted with errors.

Here, we implement the following technique to avoid these problems. Given the assumption on the wage generating process outlined in the previous section it is possible to recover the distribution of the unobserved wage component[20]:

$$\epsilon_i|ln(w_i) \textit{ not observed} \sim N(-\hat{\sigma}_\epsilon\hat{\rho}\lambda(\boldsymbol{Z}_i\hat{\alpha})\,;\, -\hat{\sigma}^2_\epsilon[1-\hat{\rho}^2\psi(\boldsymbol{Z}_i\hat{\alpha})]) \tag{3.18}$$

Where $\psi(\boldsymbol{Z}_i\hat{\alpha}) = \lambda(\boldsymbol{Z}_i\hat{\alpha})(\lambda(\boldsymbol{Z}_i\hat{\alpha}) + \boldsymbol{Z}_i\hat{\alpha})$. We use 50 random numbers for each observation from the latter distribution. Then we add this random part to the predicted mean before taking the exponential. Defining r as the r*th* draw, the predicted log-hourly gross wage for non-worker i is:

[20]See Greene (2007) for a proof.

$$ln(w_i) = \boldsymbol{X}_i\hat{\beta} + \epsilon_i^r \tag{3.19}$$

Finally, notice that:

$$exp(E[ln(x)]) \neq E[exp(ln(x))] = E[x] \tag{3.20}$$

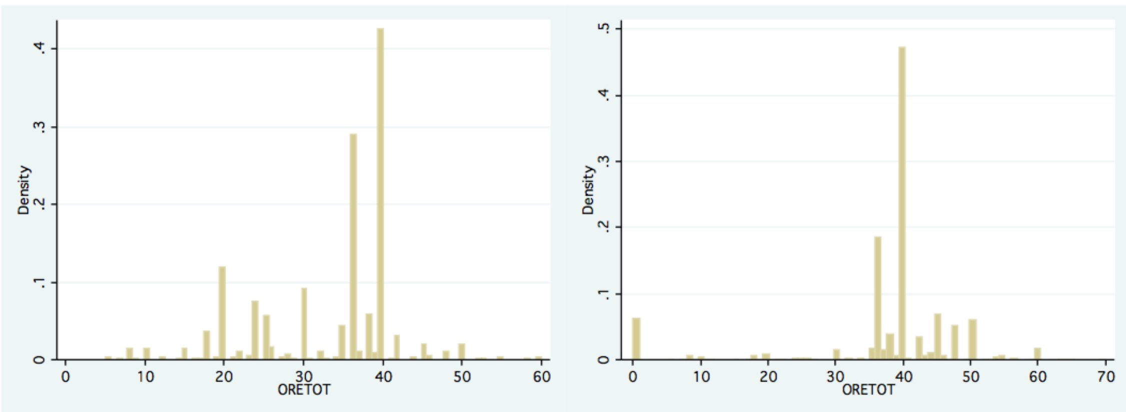
But if $ln(x) \sim N(\mu, \sigma^2)$, then:

$$E[x] = e^{\mu}e^{\frac{1}{2}\sigma^2} \tag{3.21}$$

Which in our case means:

$$E[ln(w_i)|ln(w_i)\ is\ not\ observed; \boldsymbol{X_i}] = exp(\boldsymbol{X}_i\hat{\beta}+\epsilon_i^r)exp(\frac{\hat{\sigma}_\epsilon}{2}[1-\hat{\rho}\psi(\boldsymbol{Z}_i\hat{\alpha})]) \tag{3.22}$$

This represents the expected wage for non-worker i given the r^{th} draw. Once hourly gross wages are obtained for all the observations in the relevant sample, the tax-benefit simulator is used to get the net labour incomes for each possible choice of discrete worked hours. The discrete points are defined according to the observed distribution of worked hours. The graphs below show such a distributions for both men and women in couples[21].



[21]The graphs refer to the selected sample. The graph for men includes non participation. The graph for women includes only the intensive margin (the participation rate for the selected sample of women is 49.7%).

According to these distributions, women in a couple are restricted to choose from the following discrete set: $x=\{0, 10, 20, 30, 40\}$. These points correspond to the following intervals: *0-5, 6-15, 16-25, 26-36, >36*. For married men we select the following discrete set: $y=\{0, 40, 50\}$, which corresponds to the intervals: *0-10, 11-42 and >42*. Since the labour supply for married female and men is estimated jointly, each couple has a choice set defined by the *cartesian* product $y \cdot x$ which leads to 15 possible combinations of discrete points.

The modified version of *MAPP02* computes total benefit entitlements and total tax amounts for each possible combination of discrete points given gross hourly wages. The algorithm takes time since for k possible choices of hours and *50* draws each non-worker has *k50* different net labour incomes. This means that the computational time increases exponentially when the choice set is expanded.

The net income for a particular alternative is computed by subtracting taxes and adding benefits plus non-taxable incomes to the gross labour income. This amount is then added up over the two spouses so as to get the total net household income.

Before proceeding with the estimation of the structural model, information on child-care costs and usage has to be collected from the ISTAT database. In this latter database we first drop the households without children younger than 7 years because for these households any information on child-care expenditures is recorded. Obviously, this represents a restriction due to data constraints but it must be pointed out that the school for kids is the most expensive one in Italy. Indeed, children that have turned six have access to the public school, which is basically free of charge.

Using only this sub-sample, we define 8 groups for child-care expenditure according to the presence of children aged less than 3, geographic area (northern, southern) and mother's education (low or high). For each group we compute the distribution of hourly expenditure in formal child-care for those households observed demanding formal child-care. Then we approximate this distribution using 4 mass points. Next, we compute sub-groups controlling for the mother's working status, which implies that 16 groups are now defined. For each of them we compute the percentage of households that have zero spending in child-care so that for each mother in each sub-group the probability of zero and positive spending is known.

This set of probabilities is used to integrate-out child-care quality from the likelihood function. In practise, this means that for a mother in a particular group the likelihood is evaluated 5 times (when price=0, price=quartile1, p=quartile2, etc). Then, the expected contribution to the likelihood is computed using these probabilities as weights. The expected probability of choosing alternative j for observation i is then given by:

$$Pr(p_c = 0)Pr(H_j|\boldsymbol{X}_i, \boldsymbol{u}, p_c = 0) + \sum_{s=1}^{4}[(1-Pr(p_c = 0))\cdot 0.25]Pr(H_j|\boldsymbol{X}_i, \boldsymbol{u}, p_c^s > 0) \tag{3.23}$$

Where $Pr(H_j|\boldsymbol{X}_i, \boldsymbol{u}, p_c^s)$ is the probability of choosing alternative j as defined above.

The next step is to compute the statistical relationship between hours of work and hours of child-care. This is done by running simple OLS regressions for 6 groups defined by the number of children and their age, without controlling for any sample selection bias. Following Blundell et al. (2000) we assume a linear relationship between hours of work and hours of child-care so that for each group the following child-care cost function is estimated:

$$h_{cc} = \beta_0 + \beta_1 hf_j \tag{3.24}$$

Whit this information it is possible to estimate the structural model. The estimation algorithm is implemented in Stata®. An overview of the Stata® routine used for this chapter is presented in the appendix.

3.6. Results from first-stage regressions

This section contains the set of estimates for the wage equations and child-care costs and usage. Table 1 presents the estimates of the wage equations for both female and male in couples. To identify the coefficients in the wage equations a set of instruments is used. This set includes dummies for the youngest child by age as well as the household income at zero hours of work for both the spouses. As it can be seen, all terms have the expected sign in both the selection equation and the main equation. In particular, the higher is the level of education achieved, the more likely is the participation in the labour market. The same is true if the couple does not live in southern Italy. All the coefficients in front of the instruments are significant for the female equation. In particular, female participation is lower when the couple has a child and this effect increases as the child's age decreases. As expected, these variables are less or not significant for male but it is worth noting that they have the opposite sign with respect to female. Moreover, the higher are the non-labour sources of income the lower is the probability of participation for both the spouses. Finally, it is worth pointing out that the correlation coefficient between the two error terms - *rho* - is statistically different from zero and positive in both the equations.

Table 3.1: Wage equations

	Female		Male	
	coef	z	coef	z
Log hourly gross wage:				
educ2§	0.299	4.390	0.197	5.790
educ3§	0.681	7.990	0.422	12.140
educ4§	1.087	10.190	0.825	17.510
Age†	-0.004	-0.270	0.038	2.840
Age squared†	0.021	1.020	-0.029	-1.900
Area1 : Northern§	0.244	4.080	0.262	9.600
Area2 : Middle§	0.158	2.800	0.157	4.930
Home owner§	0.125	3.410	0.111	4.900
constant	1.246	3.440	0.658	2.240
Selection equation:				
Net income at 0 hours	-0.109	-2.590	-0.218	-4.070
educ2§	0.439	4.190	0.328	2.520
educ3§	1.150	10.830	0.523	3.690
educ4§	1.870	12.460	1.354	3.420
Age†	0.095	2.560	0.190	3.380
Age squared†	-0.001	-2.860	-0.002	-3.520
Home owner§	0.424	6.120	0.360	3.200
Area1 : Northern§	0.948	12.990	0.864	5.970
Area2 : Middle§	0.708	8.140	0.811	4.840
Age youngest child : 0/2§	-0.436	-3.710	0.326	1.510
Age youngest child : 3/5§	-0.377	-3.100	0.385	1.710
Age youngest child : 6/9§	-0.330	-2.910	0.370	1.750
Age youngest child : 10/16§	-0.104	-1.310	0.243	1.890
constant	-2.074	-2.770	-3.172	-2.660
rho	0.583	4.067	0.289	1.119
sigma	0.449		0.439	
Loglikelihood	-1595		-1467	
Observations	2038		2019	
Uncensored obs.	1003		1908	

Note: Models estimated by Maximum Likelihood. † denotes that the variable is measured in terms of deviation from its mean. § denotes discrete variables. Educ$ are dummies that denote the achieved degree of education, the comparison group is the lowest level.

The next tables are related with the child-care demand. Table 2 reported below shows the distribution of child-care hourly expenditure for those households that are observed using child-care.

Table 3.2: distribution of hourly child-care cost (euro)

Groups:	qtile: 12.5	qtile: 37.5	qtile: 62.5	qtile: 87.5
No kids<=3,South Italy,Low educ	0.180	0.276	0.468	1.920
No kids<=3,South Italy,High educ	0.223	0.325	0.446	1.560
No kids<=3,North of Italy,Low educ	0.333	0.499	0.669	1.404
No kids<=3,North of Italy,High educ	0.324	0.512	0.780	1.560
Kids<=3,South Italy,Low educ	0.195	0.312	0.390	1.560
Kids<=3,South Italy,High educ	0.217	0.364	0.702	2.184
Kids<=3,North of Italy,Low educ	0.429	0.758	1.443	3.432
Kids<=3,North of Italy,High educ	0.333	0.624	0.936	3.343

Note: sample size restricted to households with children in pre-school age.

As expected, the hourly child-care cost is higher, on average, in the northern Italy and among those households with children aged less than 3 years. Table 3 shows the proportion of each group and the probability of zero spending in child-care.

Table 3.3: Summary statistics for child-care usage

Groups:	%		%	%zero_exp
No kids<=3,South Italy,Low educ	7.32	mother works	5.79	5.57
No kids<=3,South Italy,Low educ		mother not works	1.53	3.39
No kids<=3,South Italy,High educ	7.65	mother works	4.26	2.51
No kids<=3,South Italy,High educ		mother not works	3.39	1.53
No kids<=3,North of Italy,Low educ	6.39	mother works	3.88	2.62
No kids<=3,North of Italy,Low educ		mother not works	2.51	4.26
No kids<=3,North of Italy,High educ	8.2	mother works	2.62	3.88
No kids<=3,North of Italy,High educ		mother not works	5.57	5.79
Kids<=3,South Italy,Low educ	16.56	mother works	14.21	7.6
Kids<=3,South Italy,Low educ		mother not works	2.35	15.08
Kids<=3,South Italy,High educ	17.21	mother works	9.62	2.35
Kids<=3,South Italy,High educ		mother not works	7.6	8.25
Kids<=3,North of Italy,Low educ	13.99	mother works	7.6	7.6
Kids<=3,North of Italy,Low educ		mother not works	15.08	5.74
Kids<=3,North of Italy,High educ	22.68	mother works	8.25	9.62
Kids<=3,North of Italy,High educ		mother not works	5.74	14.21
Total	100		100	

Note: sample size restricted to households with children in pre-school age. %zero_exp is the proportion of households that do not use formal childcare.

The proportion of households with zero spending provides evidence that households with working mothers are more likely to buy formal childcare. Moreover, the probability of using child-care increases with the mother's level of education and it is higher when the household lives in northern Italy. Finally, Table 4 shows the results of the OLS regressions for the relationship between hours of child-care and hours of work for the mother. The results are again presented by groups.

Table 3.4: OLS regression of child-care hours on worked hours

		Coeff	St. Err	z
Hours of Child-care:				
1 kid, youngest<=3	worked hours	0.088	0.027	3.28
	cons	3.731	0.650	5.74
2 kids, youngest<=3	worked hours	0.121	0.043	2.81
	cons	11.335	0.962	11.78
2+ kids, youngest<=3	worked hours	0.428	0.091	4.72
	cons	10.735	1.723	6.23
1 kid, youngest>3	worked hours	0.088	0.052	1.70
	cons	24.708	1.270	19.45
2 kids, youngest>3	worked hours	0.185	0.053	3.50
	cons	26.724	1.141	23.43
2+ kids, youngest>3	worked hours	0.094	0.112	0.84
	cons	25.370	2.229	11.38

Note: sample size restricted to households with children in pre-school age. OLS regression by groups.

As results provided show, the increment in child-care usage for a marginal increment in the hours of work is higher when the couple has a child under three years old. This increment is higher when the household has more than two children and the youngest child is under three years old.

3.7. Results from structural model estimates

This section provides results for the structural model. The next table shows estimates for couples using 50 draws[22]. As it can be seen, most of the coefficients have the expected sign. Importantly, fixed costs of working are both positive and highly significant at standard significant levels. On average, they turned out to be about €2000 per year. Since any restriction has been imposed *a priori* on the coefficient signs, it is important to verify the coherence of the estimated preferences with respect to standard textbook economic theory. In particular, it is crucial to check if the estimated utility function is quasi-concave in income for all the observations in the sample. I made this investigation by adapting the equations 3 and 4 in Van Soest (1995). As a result I found that the utility function is quasi-concave in income for 99% of the couples in the sample.

[22]The coefficients obtained with 100 draws are not statistically significant from those obtained with 50 draws. This means that 50 draws are enough to ensure convergence.

Table 3.5: Structural model estimates for couples

	Parameter	Coef.	z		Parameter	Coef.	z
α_1 :	constant	-0.084	-7.480	β_3 :	Constant	1.386	13.260
α_2 :	constant	-0.175	-3.330		Husband's age†	0.544	2.200
α_3 :	constant	-0.384	-25.380		Husband's age sqrd†	-0.079	-2.800
α_4 :	constant	0.559	3.030		Southern Italy§	-0.248	-3.990
α_5 :	constant	-0.160	-9.100		Husband's educ. (high)§	0.011	0.210
α_6 :	constant	0.033	1.720		Number of children	0.065	1.530
β_1 :	constant	2.571	12.610		Youngest child 0-6§	0.055	0.620
	Wife's age†	-0.039	-0.450		σ_3	0.024	0.168
	Husband's age†	0.131	2.300	γ_1 :	Constant	0.027	2.455
	Southern Italy§	0.220	1.910		Big city (>40.000)	0.004	0.444
	Wife's educ. (high)§	-0.247	-2.500		Southern Italy§	0.027	3.857
	Husband's educ. (high)§	-0.016	-0.340		Youngest child 0-6§	-0.003	-0.167
	Number of children	-0.087	-1.160		Wife's age†	0.029	2.636
	Youngest child 0-6§	-0.030	-0.190	γ_2 :	Constant	0.012	12.000
	σ_1	0.157	2.432		Big city (>40.000)	0.001	1.000
β_2 :	Constant	2.112	6.430		Southern Italy§	-0.001	-0.333
	Wife's age†	0.713	3.610		Youngest child 0-6§	0.001	0.500
	Wife's age sqrd†	-0.092	-4.060		Wife's age†	0.001	1.000
	Southern Italy§	-0.189	-2.030				
	Wife's educ. (high)§	0.027	0.340				
	Number of children	-0.152	-2.650				
	Youngest child 0-6§	-0.076	-2.547				
	σ_2	0.043	0.748				
	Log-Likelihood: -3348.3188				Obs: 2002 households		

Note: model estimated by Simulated Maximum Likelihood using Halton sequences (50 draws). Weekly household income divided by 1000; Women and men's worked hours divided by 10; Random terms divided by 10; α_2 and α_3 divided by 100; α_4 divide by 1000. § denotes dummy variables and † denotes that variables are measured in terms of deviation from their means. σ_s coefficients are estimated standard deviations. The vectors $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ represent fixed costs of working, $\boldsymbol{\gamma}_2$ being the additional cost of working full time.

If we now turn on the estimated coefficients we can see that most of them are in line with standard economic guesses. Indeed, as expected, women preferences for work are decreasing with the number of dependent children, in particular when the youngest child is aged under than six. For men this pattern is exactly the opposite. Interestingly, preferences for work decrease for either spouses when they are from the southern of Italy and increase with own age at a decreasing rate. Women and men with high levels of education have increasing preferences for work. Finally, the standard deviation of the income coefficient is significantly different from zero indicating that unobserved heterogeneity in preferences for income exists in the sample. To check the ability of the model to fit the data, I computed average probabilities for each category of hours and compared with the observed frequencies. As it can be seen from the next table, the model is able to replicate observed frequencies quite well, in particular when women work more than 10 hours per week.

Table 3.6: Predicted vs observed frequencies

Hours		Observed	Predicted
Wife: 0	Husband: 0	6.06	5.78
Wife: 0	Husband: 40	33.73	34.85
Wife: 0	Husband: 50	10.66	10.07
Wife: 10	Husband: 0	0.10	0.11
Wife: 10	Husband: 40	1.50	0.92
Wife: 10	Husband: 50	0.20	0.33
Wife: 20	Husband: 0	0.65	0.89
Wife: 20	Husband: 40	7.51	7.38
Wife: 20	Husband: 50	2.60	2.38
Wife: 30	Husband: 0	0.40	0.50
Wife: 30	Husband: 40	4.55	4.69
Wife: 30	Husband: 50	1.20	1.55
Wife: 40	Husband: 0	2.85	2.79
Wife: 40	Husband: 40	21.77	21.24
Wife: 40	Husband: 50	6.21	6.53
	Tot	100	100

Once the parameters of the direct utility function are estimated, it is possible to compute the labour supply elasticities numerically following Creedy and Kalb

(2005). Firstly, gross hourly wages are increased by 1% and then a new vector of net household income for each alternative of hours is computed. Secondly, the probability of each alternative is computed for both the old and the new vector of net household income by means of the following probabilities:

$$Pr(\boldsymbol{H_j}|y^p_{ij},\boldsymbol{X_i}) = \sum_{s=1}^{5} P(p^s_c|\boldsymbol{X_i})\frac{1}{R}\sum_{r=1}^{R}\frac{exp(U(y^p_{ij}-TC^s_{ij},\boldsymbol{H_j},\boldsymbol{X_i},\boldsymbol{\nu_r}))}{\sum_{k=1}^{K} exp(U(y^p_{ik}-TC^s_{ik},\boldsymbol{H_k},\boldsymbol{X_i},\boldsymbol{\nu_r}))} \tag{3.25}$$

With *p=after, before*. These probabilities are used to compute the expected labour supply for each spouse in the couple before and after the policy change[23]:

$$E[H^s|y^p_i,X_i] = \sum_{j=1}^{J^s} Pr(\boldsymbol{H_j}|y^p_{ij},\boldsymbol{X_i})\boldsymbol{H^s_j} \tag{3.26}$$

With *s=men, women and p=after, before*. Finally, the labour supply elasticity for each spouse in the couple can be computed numerically as:

$$\varepsilon_s = \frac{E[H^s \,|\, y^{after}_i,\boldsymbol{X_i}] - E[H^s \,|\, y^{before}_i,\boldsymbol{X_i}]}{E[H^s \,|\, y^{before}_i,\boldsymbol{X_i}]}\cdot\frac{1}{0.01} \tag{3.27}$$

With *s=men, women*. The next table shows such elasticities. However, it is worth noting that such elasticities have to be interpreted carefully. They are a useful summary measure of the labour supply behaviour but it has to bear in mind that they could vary substantially depending on the initial discrete hours level and the relative change in the gross hourly wages.

[23]Since husband's earnings are on average bigger then the wife's ones, I computed two sets of elasticities derived from one percentage increase in either the woman gross wage and the man gross wage.

Table 3.7: Labour supply elasticities by individual characteristics

	Husband's gross wage +1%	Wife's gross wage +1%
All married couples	0.151	0.870
Middle/Northern Italy	0.133	0.788
Southern Italy	0.229	1.240
Couple without children	0.197	0.896
Couple with children	0.106	0.851
Youngest child <6	0.084	0.730
Youngest child>=6	0.129	0.884
Wife older than 45	-	0.989
Wife younger than 30	-	0.837
Wife with high education	-	0.621
Wife with low education	-	1.112
Husband older than 45	0.203	-
Husband younger than 30	0.148	-
Husband with high education	0.107	-
Husband with low education	0.191	-

Note: High education corresponds to secondary (5 years) or tertiary education.

As it can be seen, labour supply elasticities by household characteristics are quite in line with the expectations. Female own elasticities are on average bigger than the male's one. Moreover, elasticities are bigger in the southern Italy, for households without children and for partners without high education. The next table shows average elasticities for each decile of gross equivalent income[24]. As it was expected, elasticities are much more higher for low-income households and this is particularly true for the woman own elasticities. Finally, it can be noticed a intra-household substitution effect between income and number of working hours of the partner.

[24]Equivalent gross household income corresponds to the gross household income divided by the square root of the number of members.

Table 3.8: Average elasticities by 10 quantiles of equivalent income

	Husband's gross wage +1%		Wife's gross wage +1%	
Decile of gross income	own elasticity	partner	own elasticity	partner
1	0.26	-0.19	1.10	0.00
2	0.19	-0.12	1.02	-0.01
3	0.18	-0.04	1.00	0.00
4	0.16	-0.06	1.00	0.00
5	0.16	-0.06	0.97	0.00
6	0.13	-0.08	0.90	-0.01
7	0.13	-0.08	0.84	-0.01
8	0.12	-0.11	0.67	-0.01
9	0.11	-0.15	0.66	-0.01
10	0.02	-0.28	0.49	-0.01
Total	0.15	-0.12	0.87	-0.01

3.8. Concluding Remarks

This essay has explored the performance of the discrete approach for the estimation of labour supply elasticities for married couple using Italian data. The discrete approach has several advantages with respect to the continuous one. In particular, it easily allows for highly non-linear budget sets, non-participation, joint labour supply and endogenous child-care. Several innovations have been introduced with respect earlier Italian studies. In particular, we take into account unobserved heterogeneity in preferences, child-care expenditures, errors in wage predictions and unobserved fixed costs of working. Estimated preferences are in line with the economic theory. In particular, the marginal utility of income is positive for 99% of the sample and preferences for work decrease with the number of young children and increase with age at a decreasing rate. Elasticities are derived numerically. As expected, the average elasticity of labour supply is higher for female in couples, in particular for those who belong from low-income households. Average own elasticities are higher in southern Italy and they are lower for couples with young children.

3.9. Appendix

Overview of the STATA routine

Here we explain the routine we wrote for the optimisation algorithm. Stata® has a powerful built-in optimisation routine that is able to use, even within the same searching process, different optimisation algorithms such as Nr, Bhhh, Dfp, etc. There are three possible methods to maximise a self-written likelihood function. The first one, so-called method d0, requires the analyst to provide just the likelihood function. The second one, method d1, requires the computation of both the likelihood and the gradient. Finally, method d2, requires the computation of the likelihood, the gradient and the hessian. Whenever a piece of information is not provided, Stata computes it by numerical approximation, otherwise the algorithm just fill in the provided formula. Of course methods d2 and d1 are faster, more precise and stable but they are obviously time demanding. We chose method d0 for our program. The model to be estimated is the one described above. From a technical point of view it is a mixed conditional logit model. The difference with respect to a traditional conditional logit model is the presence of unobserved heterogeneity in preferences that has to be integrated out during the estimation process. This random terms are important because they relax the IIA assumption and give the model more reliability. Integrating out the unobserved factors (here also child-care prices and the unobserved part of wages for non-workers) produce a likelihood which is difficult to compute for the presence of a multidimensional integral. Instead of using traditional quadrature methods to approximate this integral, we follow Train (2003) and use Simulated Maximum Likelihood with Halton sequences. The Stata® command *mdraws* by Cappellari and Jenkins (2003) helps to generate the Halton Sequences from which it is easy to get the correspondent draws from the multivariate normal distribution. we call the constructed draws *random1_ r*, *random2_ r,* etc. Notice that r denotes the draw and 1,2,... identify the random terms. Formally, the utility we defined in the Stata® routine is:

$$\begin{aligned} U(y_{ij};\boldsymbol{H}_j;\boldsymbol{X_i}) \quad = \quad & \alpha_1 y_{ij}^2 + \alpha_2 hf_j^2 + \alpha_3 hm_j^2 + \\ & + \alpha_4 hf_j hm_j + \alpha_5 y_{ij} hf_j + \alpha_6 y_{ij} hm_j + \\ & + (\boldsymbol{\beta}_1 \boldsymbol{x_1'} + \nu_1) y_{ij} + (\boldsymbol{\beta}_2 \boldsymbol{x_2'} + \nu_2) hf_j + (\boldsymbol{\beta}_3 \boldsymbol{x_3'} + \nu_3) hm_j \end{aligned} \tag{3.28}$$

Where:

$$\boldsymbol{\nu} \sim N\left(\begin{pmatrix}\nu_1\\ \nu_2\\ \nu_3\end{pmatrix}, \begin{pmatrix}\sigma^2_{11} & \sigma^2_{12} & \sigma^2_{13}\\ \sigma^2_{21} & \sigma^2_{22} & \sigma^2_{23}\\ \sigma^2_{31} & \sigma^2_{32} & \sigma^2_{33}\end{pmatrix}\right) \tag{3.29}$$

And:

$$\begin{pmatrix}\nu_1\\ \nu_2\\ \nu_3\end{pmatrix} = \begin{pmatrix}c_{11} & 0 & 0\\ c_{21} & c_{22} & 0\\ c_{31} & c_{32} & c_{33}\end{pmatrix}\begin{pmatrix}random1\_r\\ random2\_r\\ random3\_r\end{pmatrix} \tag{3.30}$$

The matrix $\begin{pmatrix}c_{11} & 0 & 0\\ c_{21} & c_{22} & 0\\ c_{31} & c_{32} & c_{33}\end{pmatrix}$ is a Cholesky decomposition of the variance covariance matrix defined above[25]. Finally, bear in mind that the contribution to the Simulated Log-Likelihood for observation i is:

$$\mathcal{L}_i = \sum_{s=1}^{5} P(p_c^s|\boldsymbol{X_i})\frac{1}{R}\sum_{r=1}^{R}\prod_{j=1}^{K}\left(\frac{exp(U(y_{ij} - TC^s_{ij}, \boldsymbol{H_j}, \boldsymbol{X_i}, \boldsymbol{\nu_r}))}{\sum_{k=1}^{K} exp(U(y_{ik} - TC^s_{ik}, \boldsymbol{H_k}, \boldsymbol{X_i}, \boldsymbol{\nu_r}))}\right)^{d_{ij}} \tag{3.31}$$

In order to simplify the routine, we decided to work in a typical McFadden discrete choice environment. It means that each observation is replicated as many times as the number of alternatives she can choose from[26]. This is done by using the command *expand* after have constructed the dependent variable. The dependent variable is called *didep* and is a dummy for the observed choice among the 15 possible alternatives in the choice set[27]. The next step is to adapt the variables that change with the alternatives (net income, hours of work, fixed costs and hours of child-care) to the new McFadden environment. Below the Stata® command are reported for the steps just described.

```
***run program mdraws to get two Halton sequences per observation using primes
2,3,5:
matrix p=(2 , 3 ,5)
mdraws, neq(3) dr(50) prefix(c) burn(15) prime(p)
***get normally distributed random number using the two Halton sequences:
```

[25] A Cholesky factor is such that: $\begin{pmatrix}c_{11} & 0\\ c_{21} & c_{22}\end{pmatrix}\begin{pmatrix}c_{11} & c_{12}\\ 0 & c_{22}\end{pmatrix} = \begin{pmatrix}\sigma^2_{11} & \sigma^2_{12}\\ \sigma^2_{21} & \sigma^2_{22}\end{pmatrix}$

[26] In our database each row represents a couple so that we had to construct family variables for all the individual variables we used in the structural model.

[27] The choice set is defined by the cartesian product {0,10,20,30,40}x{0,40,50}

```
forvalues r=1/50{
gen random1_‘r’=invnorm(c1_‘r’)
gen random2_‘r’=invnorm(c2_‘r’)
gen random3_‘r’=invnorm(c3_‘r’)
}
***generate a variable for the observed choices of both the wife (hour_f) and
of the husband (hour_m):
gen Hc=1 if hour_f==0 & hour_m==0
replace Hc=2 if hour_f==0 & hour_m==40
replace Hc=3 if hour_f==0 & hour_m==50
replace Hc=4 if hour_f==10 & hour_m==0
replace Hc=5 if hour_f==10 & hour_m==40
replace Hc=6 if hour_f==10 & hour_m==50
replace Hc=7 if hour_f==20 & hour_m==0
replace Hc=8 if hour_f==20 & hour_m==40
replace Hc=9 if hour_f==20 & hour_m==50
replace Hc=10 if hour_f==30 & hour_m==0
replace Hc=11 if hour_f==30 & hour_m==40
replace Hc=12 if hour_f==30 & hour_m==50
replace Hc=13 if hour_f==40 & hour_m==0
replace Hc=14 if hour_f==40 & hour_m==40
replace Hc=15 if hour_f==40 & hour_m==50
***now expand the database to get the typical McFadden setup (15 alternatives):
gen strata=_n expand 15
sort strata
***’’respfact’’ ranks the alternatives for each observation:
gen respfact=mod(_n-1,15)+1
*** ‘‘didep’’ is the dependent variables.  It takes value one for the choosen
alternative.:
gen didep=(Hc==respfact)
***re-order the net income variable:  re-order variables for each draw j from
the wage distribution and for each alterantive (from 1 to 15) in order to get
coloumns with different draws and rows with different alternatives for each
observation:
forvalues j=1/50{
quietly{ gen y‘j’=y1_‘j’ if respfact==1
```

```
replace y`j'=y2_`j' if respfact==2
replace y`j'=y3_`j' if respfact==3
replace y`j'=y4_`j' if respfact==4
replace y`j'=y5_`j' if respfact==5
replace y`j'=y6_`j' if respfact==6
replace y`j'=y7_`j' if respfact==7
replace y`j'=y8_`j' if respfact==8
replace y`j'=y9_`j' if respfact==9
replace y`j'=y10_`j' if respfact==10
replace y`j'=y11_`j' if respfact==11
replace y`j'=y12_`j' if respfact==12
replace y`j'=y13_`j' if respfact==13
replace y`j'=y14_`j' if respfact==14
replace y`j'=y15_`j' if respfact==15
}
}
***scale the income variables:
forvalues j=1/50{
quietly{
replace y`j'=y`j'/1000
}
}
***re-order the hours variable so that each row conresponds to an alternative:
gen hf=0 if respfact==1
replace hf=0 if respfact==2
replace hf=0 if respfact==3
replace hf=10 if respfact==4
replace hf=10 if respfact==5
replace hf=10 if respfact==6
replace hf=20 if respfact==7
replace hf=20 if respfact==8
replace hf=20 if respfact==9
replace hf=30 if respfact==10
replace hf=30 if respfact==11
replace hf=30 if respfact==12
replace hf=40 if respfact==13
```

```
replace hf=40 if respfact==14
replace hf=40 if respfact==15
gen hm=0 if respfact==1
replace hm=40 if respfact==2
replace hm=50 if respfact==3
replace hm=0 if respfact==4
replace hm=40 if respfact==5
replace hm=50 if respfact==6
replace hm=0 if respfact==7
replace hm=40 if respfact==8
replace hm=50 if respfact==9
replace hm=0 if respfact==10
replace hm=40 if respfact==11
replace hm=50 if respfact==12
replace hm=0 if respfact==13
replace hm=40 if respfact==14
replace hm=50 if respfact==15
***generate interaction terms (scaled):
gen double hfhm=(hf*hm)/1000
***quadratic terms:  gen double hfsq=(hf^2)/100
gen double hmsq=(hm^2)/100
***hours constraints (fixed costs of working):
gen fc1=(Hc>=4)
gen fc2=(Hc>9)
***create interaction between fc1, fc2 and household characteristics
note:  imputs omitted for simplicity
***scale hm and hf:
replace hf=hf/10
replace hm=hm/10
***re-order the variable that indicate hours of child-care for each alternative:
gen hc=hcc0 if respfact==1
replace hc=hcc0 if respfact==2
replace hc=hcc0 if respfact==3
replace hc=hcc10 if respfact==4
replace hc=hcc10 if respfact==5
replace hc=hcc10 if respfact==6
```

```
replace hc=hcc20 if respfact==7
replace hc=hcc20 if respfact==8
replace hc=hcc20 if respfact==9
replace hc=hcc30 if respfact==10
replace hc=hcc30 if respfact==11
replace hc=hcc30 if respfact==12
replace hc=hcc40 if respfact==13
replace hc=hcc40 if respfact==14
replace hc=hcc40 if respfact==15
```

The next step is to define the maximisation algorithm. For an introduction to Stata® ML algorithm see Gould et al. (2006). Once all the temporary variables that are used in the algorithm have been defined, (see below) the variances and the covariances of the random terms have to be computed[28]. Then the covariance matrix "f" can be filled in, and the Cholesky decomposition can be computed. After that, the routine starts a double loop whose aim is adding up all the single contributions to the likelihood for any possible combination of draws from the unobserved components and from the child-care distribution. In particular, for a given draw r from the wage distribution and from the multivariate normal distributions, the contribution to the likelihood for each single observation is computed 5 times in order to integrate out child-care prices. Indeed, for a given draw r and a given child-care price, it is possible to compute the three random terms and the net income minus childcare expenditure and fixed costs. Next, the utility index and the conditional logistic probability can be filled in. This process runs 5 times, one for each mass point of the child-care cost distribution (included the point for zero costs). These values are then weighted for the probability that a particular child-care price is observed[29]. This process is carried out 50 times, which is the chosen number of draws from the wage and the random terms distribution. Given

[28] What is estimated is the logarithm of the standard deviation in order to constraint the variances to be positive. For the covariances we estimate the inverse hyperbolic tangent of the correlation terms.

[29] This is done with the command *cond(x,y,z)*. In the previous sections we explained that child-care expenditure can be considered only when the couple has a child under 7 years old. When it appends, 5 prices are available for each observation, otherwise the prices are automatically set to zero. When price is zero and the couple has a young child, the weights that are applied are different so as to take into account that the probability of zero expenditure in child-care depends on the mother's working status. As it can be seen in the routine below, if the couple does not have a young child, the command cond(-) always picks up the variable prob0, which, for these couples without young children, is fix at 0.2 (given that there are 5 choices per observation). In this way no change appends for these observations when the single contributions are added up in the subsequent line.

that for each of these draws the contribution to the likelihood is evaluated 5 times for each observation, the loop has to take into account 250 contributions for each observation at the same time[30]. Finally, these single contributions are added up into the variable L2. The last step is to pick up for each observation only the contributions that corresponds to the observed choice and to average them over the 50 draws.

```
*******************************
***Maximum Likelihood Model**
*******************************
sort strata Hc
program define clogit_sim_d0
args todo b lnf tempvar L1 L2 beta1 beta2 beta3 gamma1 gamma2 numer sum denom
tempname alpha1 alpha2 alpha3 alpha4 alpha5 alpha6 sigma1 lnsig1 l11 sigma2
lnsig2 l22 sigma3 lnsig3 l33 atrho21 atrho31 atrho32 l21 l31 l32 cov21 cov31
cov32
local d "$ML_y1"
mleval ‘alpha1’ = ‘b’, eq(1)
scalar mleval ‘alpha2’ = ‘b’, eq(2)
scalar mleval ‘alpha3’ = ‘b’, eq(3)
scalar mleval ‘alpha4’ = ‘b’, eq(4)
scalar mleval ‘alpha5’ = ‘b’, eq(5)
scalar mleval ‘alpha6’ = ‘b’, eq(6)
scalar mleval ‘beta1’ = ‘b’, eq(7)
mleval ‘beta2’ = ‘b’, eq(8)
mleval ‘beta3’ = ‘b’, eq(9)
mleval ‘gamma1’ = ‘b’, eq(10)
mleval ‘gamma2’ = ‘b’, eq(11)
mleval ‘lnsig1’ = ‘b’, eq(12) scalar
mleval ‘lnsig2’ = ‘b’, eq(13) scalar
mleval ‘lnsig3’ = ‘b’, eq(14) scalar
mleval ‘atrho21’ = ‘b’, eq(15) scalar
mleval ‘lnsig31’ = ‘b’, eq(16) scalar
mleval ‘lnsig32’ = ‘b’, eq(17) scalar
```

[30] In our PC with 4MB of RAM and 2.2Ghz Intel core duo processor, the maximisation process lasts for more than 20 hours.

```
qui gen double ‘L1’=0
qui gen double ‘L2’=0
qui gen double ‘numer’=0
qui gen double ‘sum’=0
qui gen double ‘denom’=0
scalar ‘sigma1’=(exp(‘lnsig1’))^2
scalar ‘sigma2’=(exp(‘lnsig2’))^2
scalar ‘sigma3’=(exp(‘lnsig3’))^2
scalar ‘cov21’=[exp(2*‘atrho21’)-1]/[exp(2*‘atrho21’)+1]*(exp(‘lnsig2’))*(exp(‘lnsig1’))
scalar ‘cov31’=[exp(2*‘atrho31’)-1]/[exp(2*‘atrho31’)+1]*(exp(‘lnsig3’))*(exp(‘lnsig1’))
scalar ‘cov32’=[exp(2*‘atrho32’)-1]/[exp(2*‘atrho32’)+1]*(exp(‘lnsig3’))*(exp(‘lnsig2’))
matrix f= (‘sigma1’ , ‘cov21’ , ‘cov31’ \ ‘cov21’ , ‘sigma2’ , ‘cov32’ \ ‘cov31’
, ‘cov32’ , ‘sigma3’)
capture matrix U=cholesky(f)
scalar ‘l11’=U[1,1]
scalar ‘l21’=U[2,1]
scalar ‘l31’=U[3,1]
scalar ‘l32’=U[3,2]
scalar ‘l33’=U[3,3]
forvalues r=1/50{
forvalues c=0/4{
qui gen double ‘random1’=random1_‘r’*‘l11’
qui gen double ‘random2’=random_1‘r’*‘l21’+ random2_‘r’*‘l22’
qui gen double ‘random3’=random_1‘r’*‘l31’+ random2_‘r’*‘l32’+ random_3‘r’*‘l33’
qui gen double ‘y’=y‘r’-hc*p‘c’-‘gamma1’-‘gamma2’
qui gen double ‘yhf’=‘y’*hf
qui gen double ‘yhm’=‘y’*hm
qui gen double ‘ysq’=‘y’^2
qui gen double ‘utility’=‘alpha1’*‘ysq’+‘alpha2’*hfsq+‘alpha3’*hmsq+‘alpha4’*hfhm
+‘alpha5’*‘yhf’+‘alpha6’*‘yhm’+(‘beta1’+‘random1’)*‘y’+(‘beta2’+‘random2’)*hf+
(‘beta3’+‘random3’)*hm
qui gen double ‘numer’=exp(‘utility’)
qui by strata:  gen double ‘sum’=sum(‘numer’)
qui by strata:  gen double ‘denom’=‘sum’[_N]
qui gen double ‘L1’=(‘numer’/‘denom’)*cond(p‘c’==0,prob0,prob1*0.25)
qui replace ‘L2’=‘L1’+‘L2’
```

```
drop ‘y’ ‘yhf’ ‘yhm’ ‘ysq’ ‘numer’ ‘sum’ ‘denom’ ‘L1’ ‘utility’ ‘random1’ ‘random2’
‘random3’
}
}
mlsum ‘lnf’=ln(‘L2’/50)
if ‘d’==1 if (‘todo’==0 | ‘lnf’>=.  )  exit
}
end
```

4. The recent reforms of the Italian personal income tax: distributive and efficiency effects

4.1. Introduction

The reform of the personal income tax and more generally of the tax-benefit system has recently become a very controversial topic in Italy. In the last 15 years any new government, either from the left or from the right, has started its mandate with the intention of implementing some radical reform, but it has always ended with modest adjustments with respect to the announced claims. However the tax-benefit structure has significantly changed over the last six years, possibly creating important effects in terms of redistribution and labour supply incentives.

Many studies have examined the theoretical problems inherent in the design of a consistent reform of the Italian personal income tax (for example, De Vincenti et al., 2005). Others have focused on the distributive impacts of some of the recent reforms, considering their social welfare implications and changes in effective marginal tax rates (Baldini and Bosi, 2002, Gastaldi and Liberati, 2005). The empirical papers that have considered specific cases of reform of the Italian tax-benefit system have so far made use of static microsimulation models, without consideration for possible labour supply effects, apart from calculating changes in effective marginal rates.

A given modification of the structure of marginal tax rates, however, can be followed by very different efficiency consequences, depending on the magnitude of labour supply elasticities and on their distribution across the population. Many recent papers focus on the possible equity-efficiency trade-offs implicit in any tax or benefit reform, and introduce in traditional static microsimulation models reaction functions by taxpayers and other family members. These studies have been applied to the equity and efficiency effects of the implementation, in Italy, of some basic structural reforms of the tax-benefit system (Aaberge et al., 1998), as the switch to a negative income tax (Aaberge et al., 2000) or the introduction of a guaranteed income scheme (Berliri and Parisi, 2006; Mancini, 2007).

In this paper we use a structural microsimulation model that allows to consider both the distributive and the efficiency effects of a given reform. We focus not on hypothetical general reforms of the whole system of personal taxation, but on three adjustments of the structure of the Italian personal income tax that have actually taken place in the last few years. The results show what would have happened to income distribution and labour supply if the only change in the economic environment had been represented by the modifications in the personal income tax and, in one case, in family benefits. In this sense, they do not correspond to what has actually taken place, because changes in the tax-benefit system are only one of the many factors that can influence variations in inequality and labour supply during a given period of time. These simulations capture the "pure" effects of the reforms, something that cannot be easily observed in cross-sectional data.

The model that we use integrates a detailed static tax-benefit simulator with a household labour supply model based on the Banca d'Italia survey on household income and wealth (SHIW). We study the effects of the three most recent reforms of the personal income tax on income distribution and on incentives, and show whether and in which occasions equity and efficiency moved in the same direction or followed different paths.

The next section describes the evolution of the Italian personal income tax over the past six years. In section 4 we introduce the empirical framework for our analysis. Section 5 contains the results and section 6 concludes.

4.2. The recent evolution of the Italian personal income tax

We simulate the effects of three subsequent reforms of the personal income tax (Irpef, i.e. Imposta sul reddito delle persone fisiche) that have taken place in Italy over the last few years. These changes were inspired by very different intellectual benchmarks, but none of them has come close to a complete and consistent reform. They are mainly partial attempts that have left the most important component of the Italian tax system in what is now an uncertain and unstable equilibrium.

The starting point of our analysis is the structure that Irpef had before all the reforms we simulate. Since the first of them come into force in the fiscal year 2003, we must adopt as a base case for our simulations the characteristics of Irpef in 2002. Here we sketch the main traits of the personal income tax before and after each of the reforms.

4.2.1. Irpef 2002

During the fiscal year 2002, the individual taxable income is subject to five brackets, with the lowest rate at 18% and the highest at 45% (starting from 70,000 euros). Progressivity is realised also through a series of tax credits, all piecewise decreasing with respect to income, so as to further strengthen the rise in the effective average tax rate: for dependent spouse, for children (starting at about 500 euros for each child for middle-low income levels), for dependent workers and pensioners (decreasing with respect to income from 1,150 to 52 euros), and for the self-employed (from 573 to 52 euros). The presence of these tax credits produces a minimum level of income that is exempt from the tax. This no-tax area corresponds to 6,200 euros for dependent workers and pensioners, and to 3,100 euros for independent workers.

4.2.2. Irpef 2003

The first of the two reforms operated by the centre-right government became effective in the fiscal year 2003. The number of rates was maintained at 5, but with changes in the first three rates. This reform had effects for low and middle incomes, and replaced the tax credits for earned incomes and pensions with deductions. The tax credits for dependent family members were maintained. The new deductions had therefore the main aim of guaranteeing a no-tax area, whose levels were also raised from 6,200 to 7,500 euros for dependent workers, from 6,200 to

7,000 for pensioners, and from 3,100 to 4,500 euros for the self-employed. To further enhance the progressivity effect, the tax deductions were defined as a linearly inverse function of income, falling to zero for incomes greater than 33,500 euros for dependent workers, 30,500 for the self-employed and 33,000 for pensioners. The reduction in tax revenue from these changes has been estimated in 6 billions euros.

4.2.3. Irpef 2005

The second module of reform accomplished by the centre-right government replaced also the family-related tax credits with deductions (linearly decreasing with income like the no-tax area deductions) and reduced the number of brackets from 5 to 4. The highest tax rate has been cut by two percentage points, from 45% to 43%. This top rate applies to the income share exceeding 100,000 euros. Under Irpef 2003, the 45% rate applied instead to incomes starting from 70,000 euros. While the 2003 reform benefited middle and low incomes, this reform provided tax rebates for the highest deciles of the income distribution.

The reduction of the top rate and its application to a narrower bracket were steps that the government took along a path that had, as a final objective, a structure of the personal income tax based on only two brackets: the first one up to 100,000 euros, taxed at 23%, and the rest subject to the 33% rate. Progressivity would have been further enhanced by a deduction decreasing with income. The intellectual reference point of the whole reform action of the centre-right government in this context is clearly the flat-rate tax, with a limited degree of progressivity and the application of the same legal tax rate (23%) to the great majority of taxpayers. The cost of this second piece of reform has been broadly similar to that of the first one, around 6 billion euros.

4.2.4. Irpef 2007

The flat-tax plan could not be completed because of fears of excessive revenue losses, and above all because the objective of a two-rate scheme was not shared by the centre-left government that took power in 2006. In the budget law for 2007 the new coalition introduced a deep change in the structure of the personal income tax, which is still basically effective. The top rate has been kept at 43%, but now applies to incomes above 75,000 euros. The main deductions have been replaced by tax credits, all linearly decreasing with respect to income. Formal tax rates have been reduced for middle-low incomes, but raised for those earning more than 40,000 euros (or more, if the taxpayer has dependent family members). Unlike the two previous reforms, this one accounted for a deep restructuring in cash transfers for households with children (Assegno al nucleo familiare). Before the reform, this benefit decreased in a piecewise way with respect to family income, therefore producing high marginal effective tax rates and risks of poverty traps. Now its amounts have been increased, and its structure is linearly decreasing with respect to family income.

According to official estimates, this complex reform has had a very limited cost: the reduction of the tax burden and the rise in family benefits for middle and low incomes have been financed by taxpayers with higher incomes or without children. The intellectual paradigm of the centre-left coalition was completely different from that of the preceding government: the Prodi government tried to move the first steps towards a negative income tax scheme, integrating together in a consistent scheme the personal income tax and cash transfers to families with children, so as to guarantee an income support to taxpayers with family burdens and low incomes. This objective, however, had a life even shorter than the flat-rate tax scheme, given the rapid fall of the government and its replacement with a new centre-right coalition, in power since May 2008. The new government has not introduced any relevant modification in the personal income tax so far.

4.3. Empirical Methodology and data

In order to evaluate the distributive and work incentive effects produced by the last three reforms of the Italian personal income tax, we make use of the behavioural microsimulation model introduced in chapter three. The model has two main parts: a static detailed simulator of the Italian tax-benefit system for each reform - *MAPP02* - and a microeconometric labour supply model based on utility maximising agents. The static model recovers gross earnings from net earnings provided in the survey. In this way, it is possible to compute net household income for each possible tax-scenario so as to analyse the changes in the income distribution from one reform to the other.

However, any conclusion based on the static distributional analysis is partial because it does not take into account efficiency considerations. In order to consider this latter aspect, we make use of the second part of our model that allows us to compute labour supply changes from one reform to the other. In this chapter we allow for a flexible labour supply only for couples and consider singles as being fixed on the observed labour supply behaviour.

As we have explained in chapter three, we use a unitary model of labour supply. Given the unitary framework, we consider the couple as the decision maker, which means that the two spouses choose simultaneously a combination of hours of work for each of them in order to maximise a joint utility function defined over the net household income and the hours of work of both partners.

Moreover, the labour supply model we develop is based on a discrete choice framework. In other words, we treat the number of average weekly worked hours contained in our database as a category variable and consider the couple as a utility maximising agent who chooses the combination of both worked hours and net income that gives the highest utility. As we have seen in the previous chapter, the main advantage of the discrete choice framework is its versatility in dealing with problems like non-participation and non-convex budget sets.

Our main source of data is the survey on household income and wealth (SHIW) that is conducted by the Banca d'Italia every two years. The survey collects very detailed information on income as well as social and demographic characteristics. In the present study we focus on the cross sectional survey for the year 2002. The database is representative of the whole Italian population and contains about 21,000 observations and 8,000 households.

As we have argued in the previous chapter, our model of labour supply is not appropriate to describe the labour supply decisions of any kind of household. Hence, we allow for a flexible labour supply only for a selected sub-sample of the whole population. In particular, we do not consider couples with spouses who are aged over 60 years, those who are self-employed, involved in a full time education program or serving the Army[1].

The discrete set of hours each spouse can choose from is mainly defined according to the empirical distributions observed in SHIW 2002, which are shown in picture 1 of chapter three. According to these distributions, married women are restricted to choose from the discrete set *x={0, 10, 20, 30, 40}*, while for married men we select the discrete set *y={0, 40, 50}*. Since the labour supply for married women and men is estimated jointly, each couple has a choice set defined by the *Cartesian* product *yx* that leads to *15* possible combinations of discrete points. Table 2.1 summarises the observed distribution of worked weekly hours according to these categories.

Table 4.1: Observed distribution of workers by weekly hours

Hours, men	0 (6.36%)	40 (72.22%)	50 (21.42%)	Total
Hours, women				
0	71.65	48.58	50.47	50.45
10	1.57	2.08	0.93	1.8
20	3.94	10.88	12.38	10.76
30	3.15	6.44	6.07	6.16
40	19.69	32.02	30.14	30.83
Total	100	100	100	100

Source: our computations on SHIW data. Sub-sample with flexible labour supply.

To compute net household incomes for each alternative we use our static microsimulation model. This model allows us to recover the gross hourly wages for those who are employed. Gross hourly wages correspond to gross weekly earnings from employment divided by average weekly hours of work declared in the

[1] As explained in chapter three, we make use of another database to recover information about child-care costs and child-care usage. The source of data is the survey "MULTISCOPO" 1998 on Households and Childhood Conditions that is conducted by the Italian national institute of statistics.

database. For people observed as not employed, gross hourly wages are estimated controlling for sample selection[2].

Once the hours of work have been discretised and gross wage been recovered for all the population of interest, potential gross earnings from employment for each category of worked hours are obtained by multiplying gross hourly wages by the representative hours of work in each category[3]. Then, the static tax-benefit simulator computes total benefit entitlements and total tax amounts for each potential gross annual income from employment given non-labour incomes. The net incomes are then computed by subtracting taxes and adding benefits plus non-taxable incomes to the gross labour income. This amount is then added up over the two spouses to get the total net household incomes for each alternative.

Each couple has a choice set defined over net household incomes and hours of work of either spouses and chooses the alternative that produces the maximum utility given the actual tax-benefit system and individual characteristics. Under the assumption that utility is not deterministic, it is then possible to recover the probability of a given choice, which allow estimating the preference parameters by maximum likelihood.

The labour supply model we have presented in chapter three is flexible and allows for many important extensions related to the labour supply choice of married couples. In particular, it takes into account endogenous fixed costs of working, unobserved preference heterogeneity for consumption and leisure of either spouses, endogenous child-care expenditures and errors in wage predictions for non-workers. A detailed explanation of the model was presented in the previous chapter.

[2]See the wage generating process outlined in chapter three, equations 7-9.

[3]Notice the assumption that gross hourly wages do not depend on the amount of worked hours. See Brewer et al. (2006) for this point.

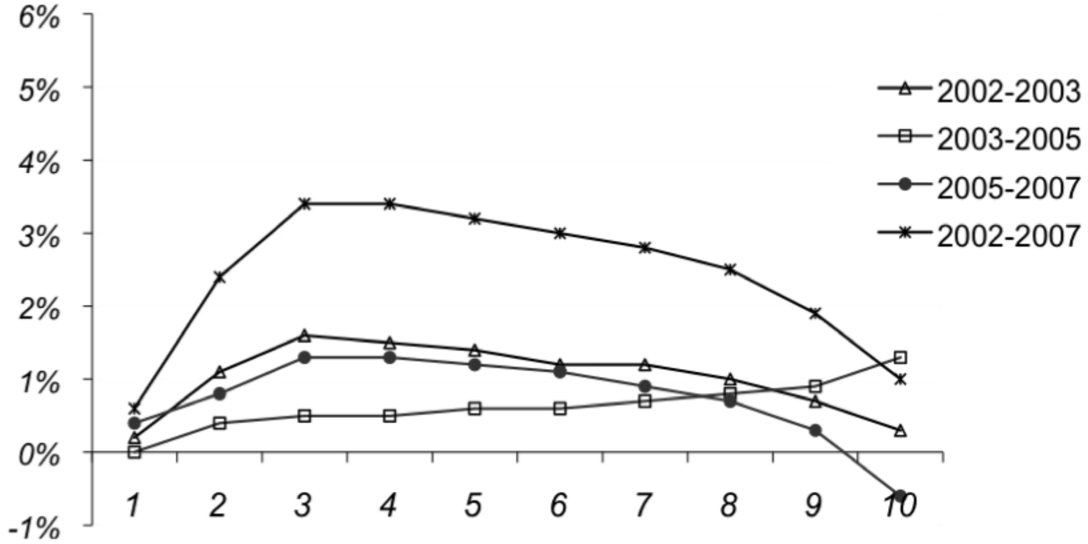
4.4. Empirical results

4.4.1. First-round distributive effects

Fig. 1a shows the effects of the various reforms on average equivalent disposable incomes of households, ordered by deciles of gross equivalent income in 2002, i.e. equivalent income before the application of the personal income tax and of family benefits[4]. In the figure we present changes in disposable equivalent incomes, due simply to the modification of fiscal parameters (in all reforms) and family benefits (only in 2007), without any behavioural reaction. The first reform, form 2002 to 2003, had a modest redistributive effect, with the third decile benefiting from the highest relative change in disposable income. The gain then declines smoothly for the richest deciles. It is very low also for the poorest 10% of households because many of them were already exempt from the personal income tax. The second reform, in 2005, has completely different distributive effects, with percentage changes in income always increasing from the poorest to the richest deciles. Finally, the adjustments introduced by the centre-left coalition in 2007 resemble those of the first centre-right module: the highest gains are achieved by the third and fourth decile. Unlike that reform, however, this episode resulted in a reduction in disposable income for the richest households.

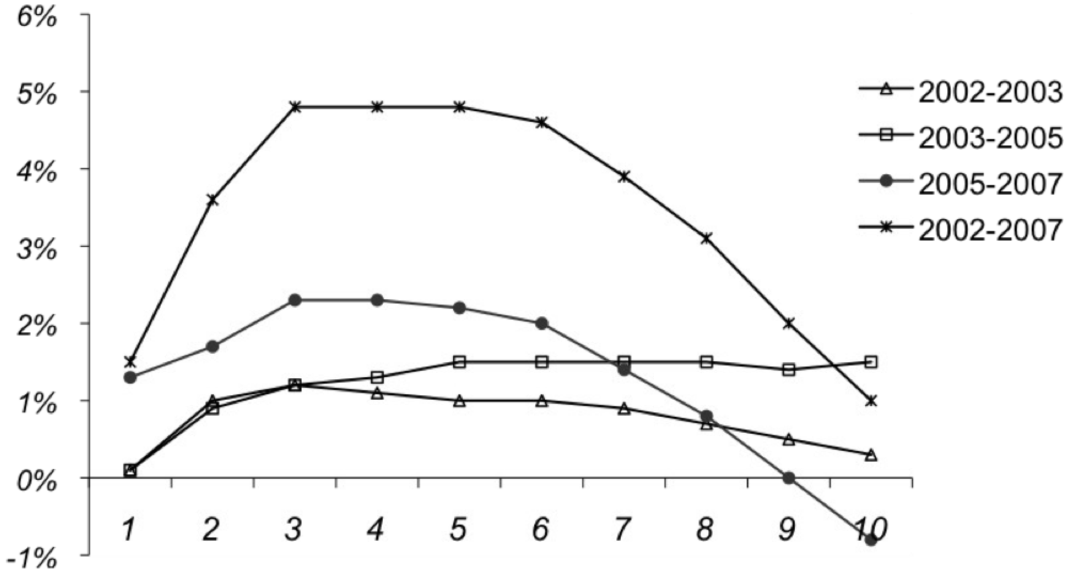
[4]Equivalent household income is household disposable income divided by the square root of the number of members.

Figure 4.1a: Percentage changes in equivalent disposable income - entire population



Focusing only on households with at least one child under 14 (Fig. 1b), the total effect of the three reforms is more generous with them than with the whole population mainly because of the 2007 reform, which turns out to have particularly benefited households with children and below median income. The income gains of the poorest households with children, however, have been very modest.

Figure 4.1b: Percentage changes in equivalent disposable income - households with at least one child under 14



Finally, Fig. 1c presents income changes only for households for whom we have simulated the possibility of changing labour supply[5]. The distributive effects are very similar to those of the other two figures.

[5] As it has been clarified in chapter three, the sub-sample with flexible labour supply corresponds to households of couples with both spouses in working age. From this sample are excluded households with self-employed spouses.

Figure 4.1c: Percentage changes in equivalent disposable income - households with flexible labour supply

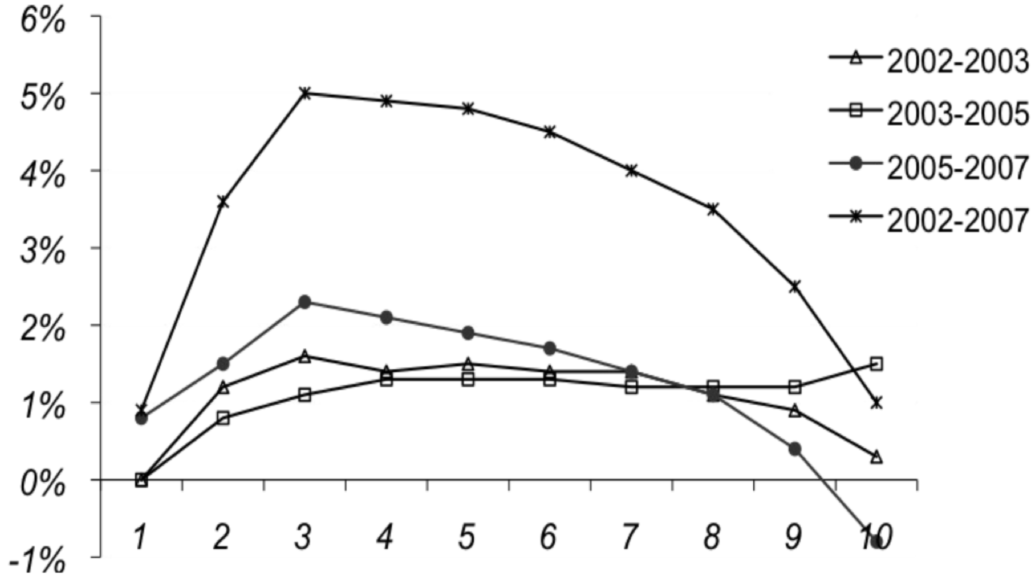


Table 2 shows the Gini indexes for disposable equivalent income, before and after each simulated reform, for the whole sample representative of the Italian population. The first reform module of 2003 reduced, as expected, inequality, but has been followed by another reform that went in the opposite direction. The centre-left reform, on the other hand, has had a more significant effect in reducing inequality. At the end of the period, inequality has been very slightly reduced: the percentage reduction from Gini 2002 to Gini 2007 is -0.94%.

Table 4.2: Gini Index and percentage change with respect to the previous year

	Gini index	% change
2002	0.35	-
2003	0.34	-0.51%
2005	0.35	0.44%
2007	0.34	-0.86%

Source: our computations on SHIW 2002 - whole sample

4.4.2. Accounting for labour supply changes

This section considers the changes in labour supply that, according to the behavioural microsimulation model, have been induced by each reform of the personal income tax. Table 5 summarises the (expected) distribution of women and men by classes of worked hours under the 2002 legislation and after each reform.

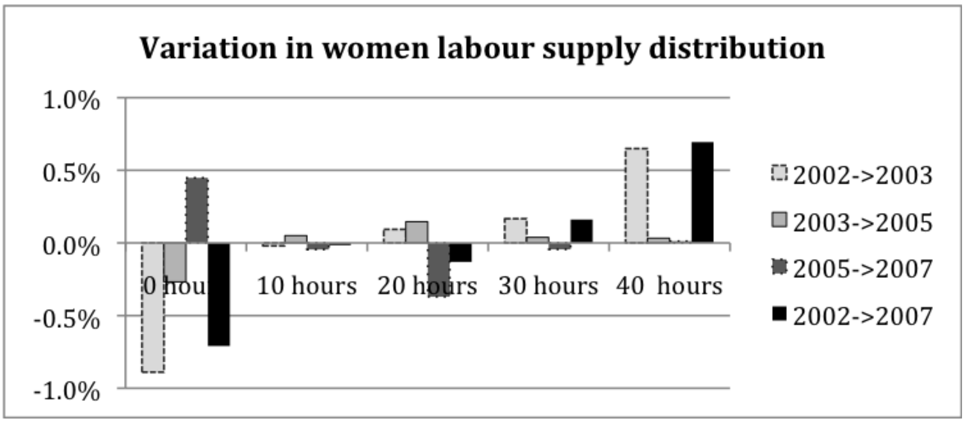
Table 4.3: Distribution of women and men by groups of weekly hours of work

Women				
	2002%	2003%	2005%	2007%
0 hours	0.51	0.5	0.49	0.5
10 hours	0.02	0.02	0.02	0.02
20 hours	0.11	0.11	0.11	0.11
30 hours	0.07	0.07	0.07	0.07
40 hours	0.31	0.31	0.31	0.31
Total	1	1	1	1
Men				
	2002%	2003%	2005%	2007%
0 hours	0.06	0.06	0.06	
40 hours	0.72	0.73	0.72	0.73
50 hours	0.21	0.21	0.22	0.22
Total	1	1	1	1

Source: our computations based on SHIW

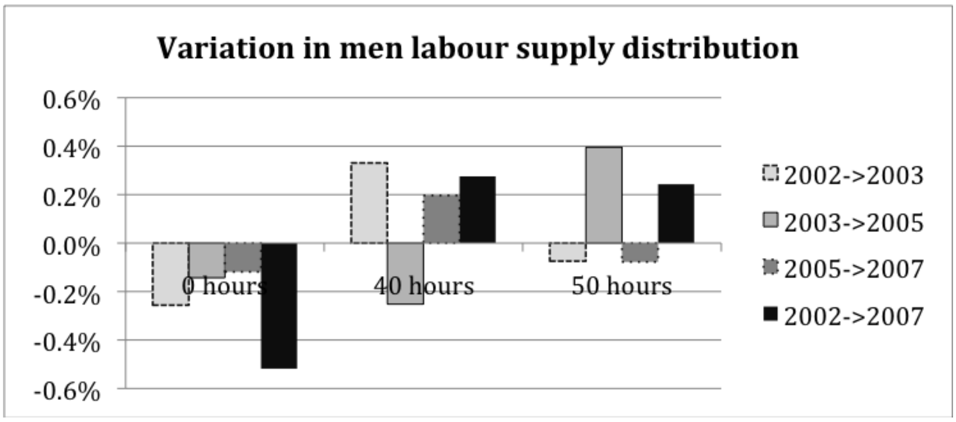
The changes induced by the modifications of PIT and child benefits are quite small. Focusing on the extensive margin, female labour supply increases after the first two reforms, but decreases slightly in 2007. Grossing up our results, we can estimate an increment of about 47,900 workers among women in couples for the 2003 reform. The 2005 module implies another increment of about 14,500 units, while the last reform reduced female participation by almost 24,000 units. During the whole period 2002-2007, women labour supply is then expected to have risen by about 38,400 units[6]. For men in couples the corresponding changes are smaller. All the three reforms increase men participation rates, in particular the last two. Grossed-up effects for men correspond to +13,000 between 2002 and 2003, +7,700 between 2003 and 2005, +6,000 between 2005 and 2007. At the end of the overall period, men's labour supply is expected to have increased by about 26,700 units due to tax and benefit reforms. If we now focus on the intensive margin, the 2003 reform increases in particular full-time jobs for women in couples. The 2005 module has effects on the number of part-time jobs instead. Finally, the 2007 reform reduces part-time jobs and does not change incentives for full-time work. For men the pattern is quite different. The only reform that actually produces changes at the intensive margin is the 2005 one. In 2007 there are no significant effects with respect to the preceding reform. Figures 2a and 2b show each variation in labour supply with respect to the previous reform for married women and men respectively. The number shown are simply the differences between the percentages contained in table 5. The graphs provide also the overall variations with respect to the baseline year.

Figure 4.2a: Changes in the distribution of workers by classes of weekly hours of work - married women



[6]It is worth stressing again that our results account for just a pure labour supply effect and cannot be compared with "real" employment data.

Figure 4.2b: Changes in the distribution of workers by classes of weekly hours of work - married men



From these two figures a complementarity between the 2003 and 2005 reforms emerges. Indeed, both reforms increase participation, but the second module of the centre-right coalition increases part-time jobs whilst the first one provides stronger incentives for full-time jobs for women. Instead, the 2007 reform reduces participation and part-time work and does not have significant effects on full time work with respect to the 2005 reform.

The second of the two graphs presents the same computations for men. As it can be seen, the changes are smaller than those for women. In general, all the reforms increase men participation with respect to the baseline year. Analysing the intensive margin, we observe opposite patterns in 2003 and 2005. In particular, the 2003 reform slightly reduces over-work incentives and stimulates full-time jobs. However, the 2005 reform increases over-work incentives to the detriment of full-time jobs. Finally, the 2007 reform acts in the same direction of the 2003 one, increasing the incentive to work full-time and reducing over-work slightly.

These findings have a clear rationale in the interaction between the reforms and the labour supply dynamics implied in our model. In particular, many simulation exercises that we do not report here have shown that the simple extension of the no-tax area has positive labour supply incentives, in particular for low-income households, whilst increasing cash benefits may result in a negative effect on labour supply[7]. Since these interventions produce possibly contrasting incentives, the labour supply pattern observed for each reform may be difficult to disentangle. However, these exercises may still help to understand the overall labour supply pattern as each reform has clear characteristics in terms of its main interventions and directions.

In particular, the 2003 reform significantly expands the no-tax area. This clearly has positive effects on participation mainly for households that have low wage rates, as an increase in labour supply may not change the tax burden. Hence, we expect the labour supply to increase for this type of households, characterised by high elasticities to marginal tax rates.

The 2005 reform reduces the tax rate for high-income households, but this turns out to have low incentive effects as high-income households have small labour supply elasticities. Hence, the main effects of the 2005 reform are expected to come from the overall change in the tax brackets and by the replacement of the family-related tax credits with deductions that linearly decrease with income.

Finally, the 2007 reform significantly increases benefits for households by means of a new system of tax-credits and cash transfers for households with dependent children. Importantly, these new tax-credits and cash transfers decrease with income at a relatively lower withdrawal rate if compared with the past schedule. Moreover, cash transfers are now significantly higher, in particular for low-income households. From our simulation exercises we found that the new structure of cash-benefits has had negative labour supply effects, whilst the extension in the withdrawal rate of the new tax-credits has ambiguous labour supply incentives as it does not automatically mean a reduction in effective marginal tax rates for all income units.

As it has been shown in the previous table, this reform has had an overall negative effect on both participation and part-time incentives, because both the child cash transfer and the tax credits for children are linearly decreasing with respect to income (household income in the case of cash transfers, individual income for tax credits), thereby raising the effective marginal tax rates to which women are exposed. This is expected to be particularly true for couples with children, as this group benefits more from this reform.

In order to check this latter point, the next table shows the differences in the labour supply distributions by classes of weekly hours worked for the sub-samples

[7]The results of these simulation exercises are available upon request.

of women with and without dependent children.

Table 4.4: Changes in the distribution of women by classes of weekly hours of work and presence of dependent children

	Couples without dependent children			*Couples with dependent children*		
Hours	var 02-03	var 03-05	var 05-07	var 02-03	var 03-05	var 05-07
0	-0.91	-0.2	0.29	-0.87	-0.43	0.6
10	-0.02	0.04	-0.04	-0.02	0.06	-0.05
20	0.12	0.15	-0.24	0.07	0.24	-0.5
30	0.17	0.03	-0.01	0.16	0.07	-0.08
40	0.63	0.01	-0.01	0.66	0.06	0.03

Source: our computations based on SHIW data

Interestingly, the variation in the labour supply distributions across hours classes is significantly different in the two sub-samples, in particular for the last reform. Indeed, comparing the variations from 2005 to 2007 in the two sub-samples we observe very different magnitudes, in particular for the extensive margin (.29 versus .6) and part-time alternatives (-.24 versus -.50). Hence, we could conclude that the 2007 reform has had a negative impact on participation incentives and on part-time jobs, in particular for women with young children.

4.4.3. Labour supply and income distribution

In this section we perform a distributive analysis of the effects of the reforms allowing for changes in labour supply behaviour, i.e. how changes in hours worked and hence in labour income have modified the distribution of net household incomes. This kind of analysis is complicated by the probabilistic nature of the labour supply model. In other words, the couples that in our database are allowed to change their labour supply have a positive probability of choosing any of the labour supply categories. Hence, if we categorise K possible alternatives of worked hours and there are N observations, we have KN possible income distributions and hence KN possible measures of the selected inequality index.

In theory, the best choice would be to consider all these possible distributions and compute a particular inequality measure KN times. Then, the correct statistic would be a weighted average of all these inequality indexes with weights equal to the probabilities of each income distribution. Of course, this approach is not practicable even with very few hour categories and few observations in the sample.

Different approaches have been proposed in the literature to solve this computational problem[8]. Here we adopt the pseudo-distribution technique proposed by Creedy et al. (2006). In practise, consider a sample of N couples allowed to have K possible labour supply alternatives. Then, we create a KN income vector with the KNth row representing the income that the N^{th} couple would have if she chose the K^{th}alternative. Each unit is weighted by the probability of observing that particular labour supply choice to create the pseudo-distribution. Creedy et al. (2006) show that standard inequality indexes computed using this pseudo-distribution converge to the real values quicker than other methods, in particular when the number of observation increases[9].

In our model there are households of singles and particular households of couples that have a fixed labour supply. In these cases, the probability attached to these records is set equal to 1. Any possible distributive analysis allowing for labour supply behaviour is implemented in this paper using this pseudo-income distribution.

Figure 3 shows the absolute variations in the labour supply distribution from one reform to the other along the various deciles of gross equivalent income and for each category of worked hours[10].

For example, the dotted line in the top-left graph shows the absolute variation in the participation rates of married women between the 2003 reform and the baseline year for each decile of equivalent income. As this line shows, married women in the first decile are those that increase participation the least with respect to the participation rate of the 2002 distribution (0.8% more). Deciles from the fourth trough the seventh one increase the participation rate the most (about 1.5% more). The continuous line in the same graph shows the variations in each decile between the participation rates implied in the 2007 reform, while the remaining line contains

[8]See Creedy et al. (2006) for a review.

[9]Convergence gets very fast when there are more than 50 observations.

[10]Deciles are computes using the pseudo-distribution methods either. This means that (equivalent) gross incomes are obtained for each alternative available to each couple and the nk vector so computed is weighted by the probability of choosing each alternative.

the variations between the participation rates of 2003 and 2005.

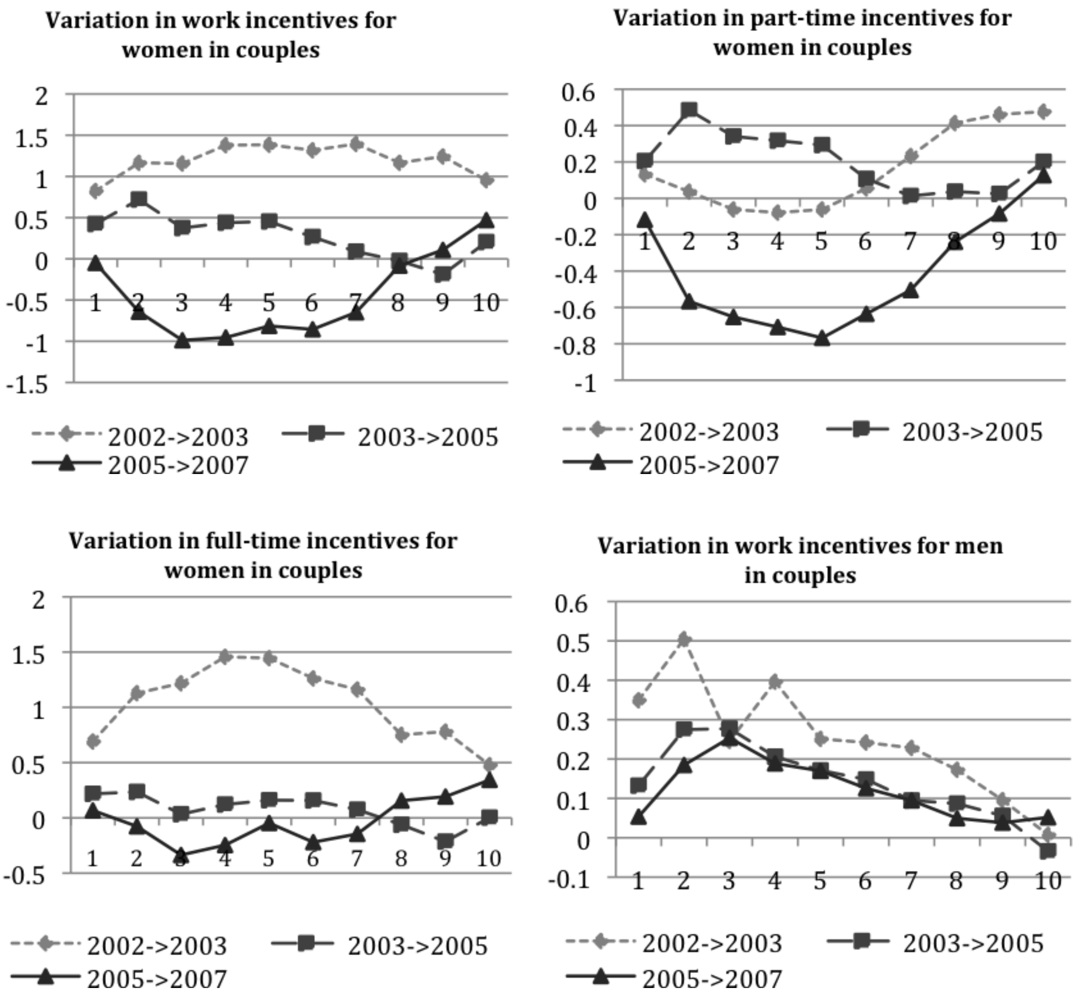
We present four graphs constructed in this way, three for women in couples and the last one for men in couples. Each graph focuses on a particular hour category. For women in couples we show results for participation (zero hours of work), part-time jobs (from 10 to 20 hours per week) and full-time jobs (from 30 to 40 hours per week) while for men in couple we present only the variations in the participation rates.

Focusing on the extensive margin for women in couples (top left graph), we can see that the 2003 and 2005 reforms have increased women participation, in particular for the middle class. The 2005 reform has had negative effects on work incentives for women in couples in the ninth decile if compared with the 2003 reform. The 2007 reform has had a negative impact on female participation, in particular for the third and fourth deciles, perhaps due to the income effect of child benefits. This latter reform has better incentives with respect to the 2005 reform for the top two deciles.

Turning to the intensive margin, the 2003 reform has strongly raised part-time jobs in the top deciles, while it has had a low negative effect for the middle deciles. The 2005 reform has had exactly the opposite pattern with respect to the 2003 one. Indeed, it increased part-time jobs for the middle class with almost no impact in the highest deciles but the very last one. The 2007 reform has had a completely different impact with respect to the 2005 reform. In particular, it has reduced part-time work incentives for almost all the deciles with the exception of the very last one. The highest reduction is registered for the third, fourth and fifth deciles. Full-time incentives for women in couples (third graph) have increased after the 2003 and 2005 reforms. Again, this is true in particular for the middle deciles. It is worth noting that the 2005 reform has not had significant effects in the top deciles and had a negative impact for the ninth one if compared with the 2003 distribution. With respect to the 2005 reform, the 2007 one has had a slightly negative impact on female full-time incentives till the seventh decile.

For men in couples (fourth graph), the analysis of the extensive margin shows that work incentives are positive for all the reforms in almost all the deciles even though the magnitudes of these incentives are smaller compared with those of women. As for women, the highest changes are registered for the middle-low income classes. Interestingly, the 2005 reform reduces men participation in the top decile with respect to the 2003 reform.

Figure 4.3: Changes in the distribution of workers by deciles of family income and types of work effort



Source: own computations based on SHIW data

We now turn to synthetic measures of income inequality to assess the overall distributive effect of each reform. The next table shows the static Gini indexes of the whole income distribution, and the percentage changes in the Gini with respect to the previous year, computed both with and without the consideration of labour supply changes.

Table 4.5. Behavioural Gini index of household disposable income before and after the reforms

Year	2002	2003	2005	2007
Static Gini	0.35	0.34	0.35	0.34
% Variation:				
Static Gini	-	-0.51%	0.44%	-0.86%
Gini with labour supply effects	-	-0.68%	0.32%	-0.91%

Source: our computations based on SHIW

The 2003 reform reduces inequality even more when labour supply is accounted for. The percentage increment of the Gini index in 2005 is smaller with respect to the static case while the 2007 reform produces a percentage reduction in inequality substantially similar to that of the static case. The principal reason of these patterns has to be found in the labour supply dynamics determined by each reform.

The 2003 reform produces greater incentives for participation and full-time jobs in the low and middle deciles. This reinforces the reduction in inequality.

After the 2005 module, we observe more increments in both participation and part-time jobs for the middle deciles; this contrasts with the rise in the Gini index due to the lower redistribution properties of the income tax.

Finally, the 2007 reform slightly reduces participation and part-time incentives for the middle deciles. However, the behavioural Gini index is marginally lower when behaviour is accounted for. A possible explanation could be found both in the reduction of part-time jobs and in the lower participation rate of women in couples. This could imply an increased homogeneity in the middle part of the income distribution that is captured by the Gini index.

The combination of income effects, homogeneity in the central section of the income distribution and pure redistributive effects explain why the static Gini index is not significantly different from the behavioural one for this reform.

4.5. Concluding Remarks

The simulation of distributive and efficiency effects of "real" reforms, when compared with the results that could be obtained working on hypothetical systemic reforms, inevitably tends to produce small changes, in particular when revenue losses are not particularly strong.

In order to make them politically acceptable, periodic adjustments to the real tax-benefit system are designed also so as to reduce the possibility that many households may lose from them. Our results actually show small changes both in the inequality measure and in labour supply, but can signal important aspects of the reforms that have been already implemented, and that have, for the simple reason of having actually taken place, a special importance in themselves.

More than on the magnitude of the behavioural changes, attention should focus on the sign of their direction. The two centre-right modules had a total cost of about 13 billion euros, with very small distributive impacts, while the subsequent reform by the centre-left government produced a greater amount of redistribution, since it actually increased the tax burden for high incomes, with no revenue loss.

From a distributive point of view, therefore, the difference between the two approaches is clear, although in all cases the changes in the Gini index have been quite modest. The adjustment in the tax structure with the most significant consequences in terms of labour supply incentives is the extension of the no-tax area in 2003. This reform produced an increase in female labour supply for low and middle deciles. As a consequence, the behavioural reactions to this reform increased real incomes at the bottom of the distribution, therefore further reducing the Gini index. This effect could not be observed using a static tax-benefit model (see fig. 1, where the effect on the first decile of the 2003 reform is negligible). The 2003 module, therefore, produced both a reduction in inequality and an increase in labour supply. In this case, we do not observe a trade-off, but complementarity, between equity and efficiency.

The 2005 module, the most apparent step towards the flat rate model, increased inequality but had a (smaller) additional positive impact on labour supply, that slightly reduced the tendency for inequality to rise. Interestingly, the behavioural contribution that in part counterbalanced the rise in inequality comes from the lower deciles since the behavioural impact of this reform in the top deciles is absolutely negligible. Indeed, husbands in the top decile reduced their labour supply while their wives slightly increased participation. The overall effect is almost zero. However, in the middle decile we observe increments in participation for both wives and husbands.

The 2007 reform, finally, has had a clear equity effect, further reducing inequality, but with a reduction in efficiency, particularly among low-income women. This could derive from the expansion of cash transfers, that are decreasing with family income and therefore produce both an income and a substitution effect on the choice between leisure and consumption, in particular for women in couples with children. The 2007 reform actually presents the traditional trade-off between efficiency and equality since it concentrates more public funds towards low and middle-income households, that have a relatively elastic labour supply. However, it is more and more difficult for reforms that come later to further improve on both the distributional and incentive effects produced by previous modifications of the tax-benefit system, in particular when the various reforms attempt to share the same broad aims, e.g. reducing inequality and/or increasing participation.

Actually, the 2007 reform preserves most of the improvements contained in the past reform as the 2005 reform did with respect to the previous one. Hence, comparing the final structure of the income tax with that of the baseline year, several steps forward emerge. Overall work incentives have improved for both women and men in couples, without reducing the redistributive effect of the personal income tax.

The broad lesson that the experience of these three reforms leaves is that it is possible to adjust the structure of the Italian tax-benefit system so as to improve both equity and efficiency. If we want to make further steps in this direction, it would be advisable to reduce marginal tax rates on low incomes, providing positive effects in terms of both income distribution and labour supply, while child benefits should not be made too rapidly decreasing with the level of family income.

5. On the role of unobserved preference heterogeneity in discrete choice models of labour supply

5.1. Introduction

As we have seen in the previous two chapters, 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 handling highly non-convex budget sets and the non-participation choice easily.

Modelling labour 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 labour 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 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

[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 labour supply models.

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

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

economic theory but on the computational problems that normally arise with gradient-based maximisation algorithms as Newton-Raphson or BHHH. Indeed, labour supply models contain a relatively high set of parameters so as to better explain how labour 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 unstable Hessian with the related probability of empirical singularity at some iterations. For this reason, the number of random parameters in labour 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 labour 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 labour supply elasticities. This finding is particularly important for the applied research whose aim is to evaluate the labour supply reaction to tax reforms empirically. Indeed, different elasticities of labour 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 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 labour 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 labour supply elasticities can lead to different results in terms of labour supply reactions and post-reform income distribution.

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 the literature. Section 3 presents an overview of the EM algorithm. Section 4 comments on the estimated utility parameters and compares elasticities across various specifications of our model. Section 5 contains the simulation and the evaluation of the introduction of a UK-style working tax-credit schedule for Italy. Section 6 concludes.

5.2. The basic econometric model without unobserved heterogeneity

In this section we develop the econometric framework for the basic structural labour supply model. Here we propose a simplified version of the basic model outlined in chapter three. In particular, we focus only on married/*de facto* couples and do not consider singles. Furthermore, we follow a unitary framework in order to model the household's decision process and assume that each household has a limited set of work alternatives. Spouses choose simultaneously the combination that maximises a joint random utility function, which is defined over the household disposable income and the hours of work of either spouse. Since the household utility is subject to a random disturbance, it is possible to recover the probability of the observed choice once an assumption on the distribution of the stochastic component is made.

In what follows, the basic model outlined in chapter three is reproposed and adapted in order to set the notation for the purposes of this chapter. In particular, 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:

$$U_{ij} = U(y_{ij},\, \boldsymbol{H_j},\, \boldsymbol{X_i}) + \xi_{ij} \tag{5.1}$$

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:

$$y_{ij} = w_{if}hf_j + w_{im}hm_j + nly_i + TB(w_{if};\, w_{im};\, \boldsymbol{H}_j;\, nly_i;\, \boldsymbol{X_i}) \tag{5.2}$$

Where w_{if} and w_{im} are the hourly gross wages from employment for women and men respectively; nly_i is the household non-labour income and the function *TB(.)* represents the tax-benefit system.

For those people who are not observed working, gross wage rates are estimated according to the standard selection model outlined in chapter three. However, in this chapter we do not allow for errors in the prediction of wages for non-workers in order to keep the model as simple as possible. For this reason, we use the estimated wages for everyone, not only for those observed out of the labour market[4].

Following the specification introduced in chapter three, 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:

$$\begin{aligned} U(y_{ij}; \boldsymbol{H}_j; \boldsymbol{X_i}) &= \alpha_1 y_{ij}^2 + \alpha_2 hf_j^2 + \alpha_3 hm_j^2 + \\ &\quad + \alpha_4 hf_j hm_j + \alpha_5 y_{ij} hf_j + \alpha_6 y_{ij} hm_j + \\ &\quad + \beta_1 y_{ij} + \beta_2 hf_j + \beta_3 hm_j \end{aligned} \tag{5.3}$$

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

$$\beta_j = \sum_{i=1}^{K_j} \beta_{ij} x_{ij} \qquad j \epsilon \{1, 2, 3\} \tag{5.4}$$

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[5]:

$$\begin{aligned} Pr(\boldsymbol{H} = \boldsymbol{H_j} | \boldsymbol{X_i}) &= Pr[U_{ij} > U_{is}, \forall s \neq j] \\ &= \frac{exp(U(y_{ij},\, \boldsymbol{H_j},\, \boldsymbol{X_i}))}{\sum_{k=1}^{K} exp(U(y_{ik},\, \boldsymbol{H_k},\, \boldsymbol{X_i}))} \end{aligned} \tag{5.5}$$

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

$$LL = \sum_{i=1}^{N} log \prod_{j=1}^{J} (Pr(\boldsymbol{H} = \boldsymbol{H_j} \mid \boldsymbol{X_i}))^{d_{ij}} \tag{5.6}$$

[4] It is worth noting that this is necessary in order to avoid biased simulation results. Indeed, using predicted wages only for non-workers without controlling for errors in the prediction, creates a distribution of wages for this sub-sample that would be too concentrated on the mean (expected) value. This, of course, has a strong impact on the subsequent variance of the simulated net-income alternatives and, in turns, on the estimation results.

If predicted wages are used for the whole population, instead, both workers and non-workers have an estimated hourly productivity level that is computed according to the same wage generating process, which allows for a more consistent analysis and for more comparable results across different sub-groups of the whole population.

[5] See McFadden (1973)

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 labour supply framework[6].

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.

[6]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.

5.3. 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[7]. 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 a great freedom in the choice of different densities and many alternatives can be tested[8].

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 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 labour 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:

$$\mathbf{U}(\mathrm{y_{ij}}, \boldsymbol{H_j}, \boldsymbol{X_i}) = \boldsymbol{W}^{'}_{\mathbf{ij}}\boldsymbol{\alpha} + \boldsymbol{G}^{'}_{ij}\boldsymbol{\beta} + \xi_{ij} \tag{5.7}$$

[7] See McFadden and Train (2000).
[8] Common choices are the Gaussian, the log-normal or the triangular distribution.

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 labour supply literature, assume now the set of parameters in vector $\boldsymbol{\beta}$ to be random:

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\Theta}\boldsymbol{X}_i + \boldsymbol{\Omega}\boldsymbol{\vartheta}_i \qquad E(\boldsymbol{\vartheta}_i) = \mathbf{0},\ Cov(\boldsymbol{\vartheta}_i) = \boldsymbol{\Sigma} \tag{5.8}$$

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:

$$Pr(\boldsymbol{H_i} = \boldsymbol{H_{ij}} \,|\, \boldsymbol{X_i}) = \int Pr(\boldsymbol{H_i} = \boldsymbol{H_{i,j}} \,|\, \boldsymbol{X_i}, \boldsymbol{\vartheta_i})\boldsymbol{\phi(\vartheta_i)}d\boldsymbol{\vartheta_i} \tag{5.9}$$

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:

$$LL = \sum_{i=1}^{N} log\frac{1}{R}\sum_{r=1}^{R}\prod_{j=1}^{J}[Pr(H_i = H_{ij} \,|\, \boldsymbol{X_i}, \boldsymbol{\vartheta_{ir}})]^{d_{ij}} \tag{5.10}$$

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 labour 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:

$$Pr(\boldsymbol{H_i} = \boldsymbol{H_{ij}} \mid \boldsymbol{X_i}, \boldsymbol{\beta_c}) = \frac{exp(\boldsymbol{W}_{ij}^{'}\boldsymbol{\alpha} + \boldsymbol{G}_{ij}^{'}\boldsymbol{\beta_c})}{\sum_{k=1}^{K} exp(\boldsymbol{W}_{ik}^{'}\boldsymbol{\alpha} + \boldsymbol{G}_{ik}^{'}\boldsymbol{\beta_c})} \tag{5.11}$$

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[9]:

$$Pr(class_i = c \mid \boldsymbol{\Delta_i}) = \frac{exp(\boldsymbol{\Delta_i^{'}}\boldsymbol{\gamma_c})}{\sum_{c=1}^{C} exp(\boldsymbol{\Delta_i^{'}}\boldsymbol{\gamma_c})} \qquad c = 1, 2, ..., C \; ; \; \boldsymbol{\gamma_C} = \boldsymbol{0} \tag{5.12}$$

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[10].

AsRoeder 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 labour 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

[9]See Greene (2001).

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

household i chooses alternative j is:

$$\sum_{c=1}^{C} Pr(class_i = c \mid \boldsymbol{\Delta_i}) Pr(\boldsymbol{H_i} = \boldsymbol{H_{ij}} \mid \boldsymbol{X_i}, \boldsymbol{\beta_c}) \tag{5.13}$$

Hence, the likelihood for the whole sample is:

$$LL = \sum_{i=1}^{N} log \sum_{c=1}^{C} Pr(class_i = c \mid \boldsymbol{\Delta_i}) \prod_{j=1}^{J} [Pr(\boldsymbol{H_i} = \boldsymbol{H_{ij}} \mid \boldsymbol{X_i}, \boldsymbol{\beta_c})]^{d_{ij}} \tag{5.14}$$

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 labour 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[11].

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 labour supply literature has always been limited and this could partially justify Haan's claim, who has not found significant differences in the labour 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 labour 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

[11] Interestingly, as we have seen with the two mixed models, the set of parameters that are traditionally assumed to be random in the labour 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.

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.

5.4. An EM recursion for discrete choice models of labour 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[12]. 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 "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 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 maximised by standard gradient-based methods as Newton Raphson or BHHH. However, it can be shown that the same log-likelihood can be maximised by repeatedly updating the following recursion:

$$\begin{aligned} \boldsymbol{\eta}^{s+1} &= argmax_{\eta} \sum_i \sum_c C_i(\boldsymbol{\eta}^s) ln\, w_{ic}(\boldsymbol{\gamma_c}) \prod_j \left[P(\boldsymbol{H_{ij}}|\boldsymbol{X_i}, \boldsymbol{\pi_c})\right]^{d_{ij}} \\ &= argmax_{\eta} \sum_i \sum_c C_i(\boldsymbol{\eta}^s) ln(L_i \,|\, class_i = c) \end{aligned} \tag{5.15}$$

Where $\boldsymbol{\pi_c} = (\boldsymbol{\beta_c}\,\boldsymbol{;}\ \boldsymbol{\alpha_c})'$, $\boldsymbol{\eta} = (\boldsymbol{\pi_c};\ \boldsymbol{w_c};\ \boldsymbol{\gamma_c},\ c = 1, 2, .., C)$, $w_{ic}(\boldsymbol{\gamma_c})$ is the unconditional density of the missing data computed as in eq. 5.12, L_i is the *joint* likelihood of both the observed choice and the missing data 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

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

means of Bayes' theorem:

$$C_i(\boldsymbol{\eta}^s) = \frac{L_i|class_i = c}{\sum_{c=1}^{C} L_i|class_i = c} \tag{5.16}$$

Now, given that:

$$ln\, w_c(\boldsymbol{\gamma_c})P(\boldsymbol{H_{ij}}|\boldsymbol{X_i,\pi_c}) = ln\, w_c(\boldsymbol{\gamma_c}) + ln\, P(\boldsymbol{H_{ij}}|\boldsymbol{X_i,\pi_c}) \tag{5.17}$$

the recursion in eq. 5.15 can be split into different steps:

1. Form the contribution to the likelihood $(L_i \,|\, class_i = c)$ as defined in eq. 5.14 *for each class*[13],
2. Form the *individual-specific* posterior probabilities of class membership using eq. 5.16,
3. *For each class*, maximise the *weighted* log-likelihood so as to get a new set of $\pi_c,\ c = 1, ..., C$:

$$\pi_c^{s+1} = argmax_{\boldsymbol{\pi}} \sum_i C(\boldsymbol{\eta}^s) ln \prod_j \left[P(\boldsymbol{H_{ij}}|\boldsymbol{X_i,\pi_c})\right]^{d_{ij}} \tag{5.18}$$

4. Following eq. 5.17, maximise the other part of the log-likelihood in eq. 5.14 and get a new set of w_c, $c = 1, ..., C$:

$$w_{ic}^{s+1} = argmax_{\boldsymbol{w}} \sum_{i=1}^{N} \sum_{c=1}^{C} C_i(\boldsymbol{\eta}^s) ln\, w_{ic}(\boldsymbol{\gamma_c}) \tag{5.19}$$

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

$$\boldsymbol{\gamma}^{s+1} = argmax_{\boldsymbol{\gamma}} \sum_{i=1}^{N} \sum_{c=1}^{C} C_i(\boldsymbol{\eta}^s) ln \frac{exp(\boldsymbol{\Delta_i'}\gamma_{\boldsymbol{c}})}{\sum_c exp(\boldsymbol{\Delta_i'}\gamma_{\boldsymbol{c}})}, \quad \boldsymbol{\gamma_C = 0} \tag{5.20}$$

[13] 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.

- and update $w_{ic}(\boldsymbol{\gamma_c})$, $c = 1, ..., C$ as:

$$w_{ic}^{s+1} = \frac{exp(\boldsymbol{\Delta_i'}\boldsymbol{\hat{\gamma}_c^{s+1}})}{\sum_c exp(\boldsymbol{\Delta_i'}\boldsymbol{\hat{\gamma}_c^{s+1}})}, \qquad c = 1, 2, ..., C \; ; \; \boldsymbol{\gamma_C = 0} \qquad (5.21)$$

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[14].

It is worth noting that in each maximisation, 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 maximisation of the whole model as in eq. 5.14.

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[15].

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 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[16].

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

$$w_c^{s+1} = \frac{\sum_i C_i(\boldsymbol{\eta}^{s+1})}{\sum_i \sum_c C_i(\boldsymbol{\eta}^{s+1})}, \; c = 1, ..., C \qquad (5.22)$$

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.

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

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

5.5. 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 labour 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 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-time and overwork), which implies 18 different combinations for each household[17]. 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)[18].

In what follows, we first consider the three models introduced in sections 1 and 2. The first model is estimated without accounting for unobserved heterogeneity and is then a typical multinomial conditional logit (MNL), as we explained in section 1.

The second model is by far the most common in the applied labour 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 labour supply modelling, we allow the three coefficients of the linear terms of the utility to be random with independent normal densities[19]. We then estimate the means and the standard deviations of these coefficients along with the other preference parameters using Simulated Maximum Likelihood[20].

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 nonparamet-

[17]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.
[18]See Baldini and Ciani (2009).
[19]The estimation with correlated normal densities did not improve the likelihood and the estimated correlation coefficients were not significant.
[20]See Train (2003).

ric estimation of mixed logit models *á 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, a primary goal is the definition of the proper number of latent classes. However, as we explained in section 2, due to the computational difficulties related to standard optimisation methods, labour 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[21].The next table shows the estimated parameters for these three models, along with the maximised log-likelihood:

[21] 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.

Table 3.1: Estimated utility parameters (1)

		MNL		RCML		HSML	
		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	1.86
	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. 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.) and 1(wife=0 ho.) are dummy for the alternatives where the husband and the wife do not work.

As the table shows, most coefficients have the expected sign over the three specifications[22]. 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[23].

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. 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 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 generalises 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[24]. 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. FollowingGreene

[22] 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.

[23] 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.

[24] 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.

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 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 2 in the appendix shows the (rotated) factor loadings obtained with the varimax rotation whose eigenvalues were higher than one[25]. Following Thompson and Daniel (1996), 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.

The next table reports the coefficients for the LCML model with four latent classes along with their (weighted) average across the four classes[26]. As can be seen, the maximised log-likelihood is significantly higher with respect to the other models and also the fitting significantly increases[27]. 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 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.

[25] 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.

[26] 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.

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

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[28].

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

Table 3.2: 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

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 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 labour supply behaviour implied in our estimated model and can be used to check whether different specifications lead to different policy prescriptions[29].

labour 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 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 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 labour supply elasticities, we adopt the same strategy as Haan (2006). More specifically, we computed 95% bootstrapped confidence intervals for the MNL labour supply elasticities and checked whether they 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 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.

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

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 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 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 A3.3: Labour supply elasticities for married couples

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

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

These findings are relevant in particular for the applied literature. Indeed, discrete choice labour supply models have been estimated only using the RCML or the HSML so far and the estimated coefficients are then used to analyse the labour 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 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.

5.6. Simulating a WTC for Italy

The aim of working-tax credits is to encourage the participation of low income households in the labour 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[30]. 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 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

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

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 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. 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 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 labour 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 labour supply pattern after the reform. However, according to the elasticities computed in the previous section, the labour supply reaction produced by the LCML model is significantly stronger with respect to the other specifications.

Table 3.4: Labour 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)

In order to better understand the differences between 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, 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.

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

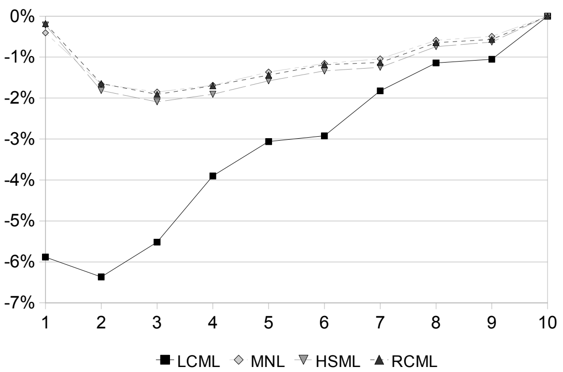


Figure 2: variation in women part time jobs for decile of gross household income

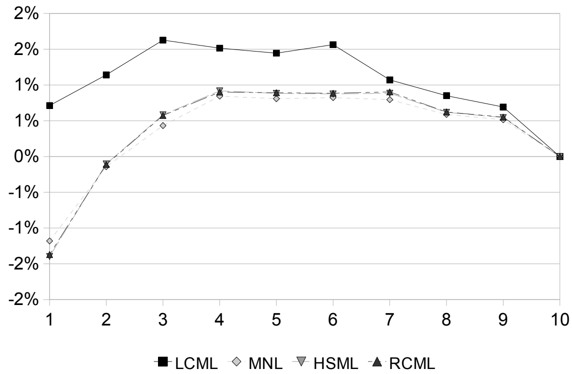


Figure 3: variation in women full time jobs for decile of gross household income

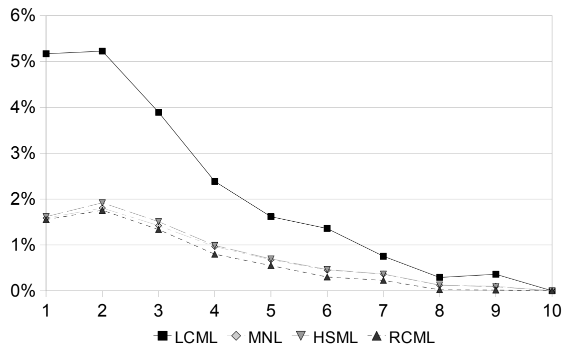
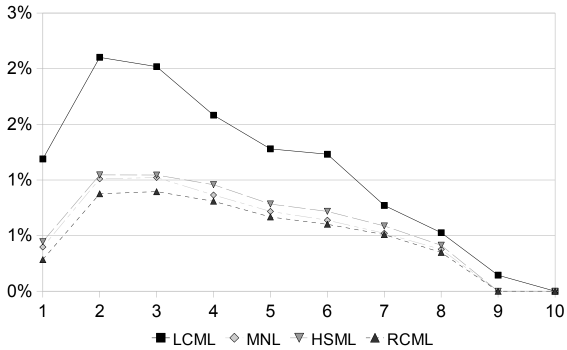


Figure 4: variation in men participation rates for decile of gross household income



In order to evaluate how the income distribution changes after the reform, we compute the Gini index before 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% versus an average of -0.84 over the other three specifications).

Table 3.5: Gini index before and after the reform

	LCML	MNL	HSML	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).

5.7. Concluding Remarks

The aim of this chapter has been twofold. First, we have shown that the way researchers account for unobserved heterogeneity can have an impact on the derived labour 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 maximisation algorithms.

5.8. Appendix (1)

Table A3.1: 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 A3.2: 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 A3.3: 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)

5.9. Appendix (2)

The aim of this appendix is to show the practical implementation of the EM algorithm for the estimation of the latent class model we proposed in the chapter. Importantly, in what follows we allow for panel data, i.e. the agent is observed making a given choice over a panel of T periods (the "choice situations"). We coded the algorithm in Stata® 10.1 using a simple *do* file. Work has been in progress in order to generalise the *do* file by writing a *ado* file. To start with, the log-likelihood we want to maximise is the following:

$$LL = \sum_{i=1}^{N} log \sum_{c=1}^{C} Pr(class_i = c \,|\, \boldsymbol{\Delta_i}) \prod_{t=1}^{T} \prod_{j=1}^{J} [Pr(\boldsymbol{H_i} = \boldsymbol{H_{ijt}} \,|\, \boldsymbol{X_{ijt}}, \boldsymbol{\pi_c})]^{d_{ijt}} \quad (5.23)$$

Where c is the number of latent classes; d_{ij} is a dummy that selects the observed choice;

$$Pr(class_i = c \,|\, \boldsymbol{\Delta_i}) = \frac{exp(\boldsymbol{\Delta_i'}\boldsymbol{\gamma_c})}{\sum_c exp(\boldsymbol{\Delta_i'}\boldsymbol{\gamma_c})}, \quad c = 1, ..., C, \quad \gamma_C = 0 \quad (5.24)$$

is the *unconditional* probability of class membership, which may depend on a set of demographics included in matrix $\boldsymbol{\Delta}_i$, and

$$Pr(\boldsymbol{H_i} = \boldsymbol{H_{ijt}} \,|\, \boldsymbol{X_{ijt}}, \boldsymbol{\pi_c}) = \frac{exp(\boldsymbol{X}_{ijt}'\boldsymbol{\pi_c})}{\sum_{k=1}^{K} exp(\boldsymbol{X}_{ikt}'\boldsymbol{\pi_c})} \quad (5.25)$$

is the probability that agent i chooses alternative j given that he/she belongs from class c. This model has been traditionally estimated by maximum likelihood.

However, as we have seen in the main text, such estimation is often difficult, in particular if the number of parameters and the number of classes are high. For these reasons, in the labour supply literature only a very small set of parameters (one up to three) is assumed to be different in each class. Moreover, the covariates that enter the probability of class membership often collapse to a simple constant and the number of classes rarely are more than two. Obviously, these constraints may affect post-estimation results and do not actually provide a fully non-parametric estimation method, as the number of classes is constrained by the researcher to a very small number.

An Em algorithm, thanks to its overall stability, may provide a better alternative for a fully nonparametric estimation of this model without these computational constraints. In particular, we have seen that the log-likelihood in equation 5.23 can be maximised by repeatedly updating the following recursion:

$$\begin{aligned} \boldsymbol{\eta}^{s+1} &= argmax_\eta \sum_i \sum_c C_i(\boldsymbol{\eta}^s) ln\, w_{ic}(\boldsymbol{\gamma_c}) \prod_t \prod_j \left[P(\boldsymbol{H_{ijt}}|\boldsymbol{X_{ijt}},\boldsymbol{\pi_c})\right]^{d_{ijt}} \\ &= argmax_\eta \sum_i \sum_c C_i(\boldsymbol{\eta}^s) ln(L_i \,|\, class_i = c) \end{aligned} \tag{5.26}$$

Where $\boldsymbol{\eta} = (\boldsymbol{\pi_c};\, \boldsymbol{w_c};\, \boldsymbol{\gamma_c},\, c = 1, 2, .., C)$; $w_{ic}(\boldsymbol{\gamma_c})$ is the *unconditional* density of class membership, which is computed as in equation 5.24; L_i is the *joint* likelihood of both the observed and missing data (that is, the class membership); $C(\boldsymbol{\eta}^s)$ is the *posterior* probability that household i belongs to class c, conditional on the density of the observed and missing data *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:

$$C_i(\boldsymbol{\eta}^s) = \frac{L_i|class_i = c}{\sum_{c=1}^{C} L_i|class_i = c} \tag{5.27}$$

Now, given the basic fact that:

$$ln\, w_c(\boldsymbol{\gamma_c}) P(\boldsymbol{H_{ijt}}|\boldsymbol{X_{ijt}},\boldsymbol{\pi_c}) = ln\, w_c(\boldsymbol{\gamma_c}) + ln\, P(\boldsymbol{H_{ijt}}|\boldsymbol{X_{ijt}},\boldsymbol{\pi_c}) \tag{5.28}$$

the recursion in equation 5.26 can be split into different steps:

1. Form the contribution to the likelihood $(L_i \,|\, class_i = c)$ as defined in equation 5.23 *for each class*[31],

2. Form the *individual-specific* posterior probabilities of class membership using equation 5.27,

3. *For each class*, maximise the *weighted* log-likelihood so as to get a new set of $\pi_c,\ c = 1, ..., C$:

$$\pi_c^{s+1} = argmax_{\boldsymbol{\pi}} \sum_i C(\boldsymbol{\eta}^s) ln \prod_t \prod_j \left[P(\boldsymbol{H_{ijt}}|\boldsymbol{X_{ijt}},\boldsymbol{\pi_c})\right]^{d_{ijt}} \tag{5.29}$$

[31] 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.

4. Following equation 5.28, maximise the other part of the log-likelihood in equation 5.26 and get a new set of w_c, $c = 1, ..., C$:

$$w_{ic}^{s+1} = argmax_{\boldsymbol{w}} \sum_{i=1}^{N} \sum_{c=1}^{C} C_i(\boldsymbol{\eta}^s) ln\, w_{ic}(\boldsymbol{\gamma_c}) \tag{5.30}$$

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

$$\boldsymbol{\gamma}^{s+1} = argmax_{\boldsymbol{\gamma}} \sum_{i=1}^{N} \sum_{c=1}^{C} C_i(\boldsymbol{\eta}^s) ln \frac{exp(\boldsymbol{\Delta_i'\gamma_c})}{\sum_c exp(\boldsymbol{\Delta_i'\gamma_c})}, \quad \boldsymbol{\gamma_C = 0} \tag{5.31}$$

- and update $w_{ic}(\boldsymbol{\gamma_c})$, $c = 1, ..., C$ as:

$$w_{ic}^{s+1} = \frac{exp(\boldsymbol{\Delta_i'\hat{\gamma}_c^{s+1}})}{\sum_c exp(\boldsymbol{\Delta_i'\hat{\gamma}_c^{s+1}})}, \qquad c = 1, 2, ..., C \ ; \ \boldsymbol{\gamma_C = 0} \tag{5.32}$$

In what follows we show how to implement this algorithm in a Stata *do* file. Importantly, it has to be borne in mind that the database is already organised for the estimation of standard conditional multinomial logit models in Stata (i.e. the command "*CLOGIT*"), which means that each row corresponds to a particular alternative for a given individual.

The first relevant point for the implementation of the EM algorithm is to define the variables (the dependent variable "*$depvar*", which has to be constructed according to the estimation of a standard *CLOGIT* model; the list of covariates that enter the probability of the observed choice "*$varlist*"; the variable that ranks the alternatives for each agent "*$alt*"; the variable that identifies the panel dimension, i.e. the choice makers "*$ind*"; the variable that defines the choice situations for each choice maker "*$idt*") and the environment (the number of classes "*$nclasses*" and the number of iterations "*$niter*"), and to provide Stata with the starting values. In order to perform this latter task, we split the sample into C different sub-samples (one for each class) and estimate a separate *CLOGIT* for each sub-sample. As for the starting values for the probability of class-membership (equation 5.24) we simply define equal shares, that is $\frac{1}{C}$:

```
  ********************************************
***define global variables and environment**
********************************************
global depvar "ch"
global varlist "introduce covariates here"
global alternative "alt"
global id "id"
**note $id defines the panel dimension (the choice makers)
global idt "ind"
**note:  $idt defines the choice situations for each choice maker
global nclasses "4"
global niter "200"
```

Importantly, EM algorithms have been proved to be very slow and they may converge to local maxima. Hence, we suggest to set a (relative) large number of iterations and to try with different starting values (the routine already computes random starting values, see below):

```
*************************
***Get starting values***
*************************
**Partition the sample into $nclasses subsamples:
sort $id $idt $alt
by $id:  gen double p=uniform() if _n==_N
by $id:  egen double pr=sum(p)
global prop 1/$nclasses
gen double ss=1 if pr<=$prop
forvalues s=2/$nclasses{
replace ss=‘s’ if pr>(‘s’-1)*$prop & pr<=‘s’*$prop
}
drop p pr ta ss
```

The next step is to estimate a separate *CLOGIT* for each sub-sample; after each estimation, use the *predict* command to get the probability of choosing each alternative in each of the *\$nclasses* latent classes; call these vectors of probabilities as *l_, l_2,...,l_C*:

```
forvalues s=1/$nclasses{
clogit $depvar $varlist if ss==‘s’, group($idt)
predict double l_‘s’
}
```

**Finally, define equal shares for the starting values:

```
forvalues s=1/$nclasses{
gen double prob‘s’=$prop
}
```

We now show the steps needed to calculate the posterior probability as in equation 5.27, given the starting values we have just computed.

Firstly, we multiply *l_1, l_2,..,l_C* by the dummy variable that identifies the observed choice in each choice situation so as to pick only the probability of the observed choice.

Secondly, for each choice maker, we multiply the probabilities of the observed choices *in each choice situation*. Importantly, notice that this last point is performed through the program *gprod*, which has to be downloaded from the Internet (type: "*findit gprod*" in the Stata command for more information).

```
forvalues s=1/$nclasses{
qui gen double kbb‘s’=l_‘s’*$depvar
qui recode kbb‘s’ 0=.
by $id:  egen double kbbb‘s’=prod(kbb‘s’)
by $id:  replace kbbb‘s’=.  if _n!=_N
}
```

The third step is to construct the denominator of eq. 5.27 by computing a weighed sum of the previous variables (*kbbb1, kbbb2, etc.*) with weights given by the probability of class membership (*prob1, prob2,..., etc.*):

```
gen double den=prob1*kbbb1
forvalues s=2/$nclasses{
replace den=den+prob‘s’*kbbb‘s’ }
```

Then we construct the ratio es defined in eq. 5.27 as:

```
forvalues s=1/$nclasses{
gen double h‘s’=(prob‘s’*kbbb‘s’)/den
qui recode h‘s’ .=0 }
```

Finally, we simply rearrange the previous variables (*h1, h2, etc.*) in order to create individual-level variables. These are the conditional probabilities $C(\boldsymbol{\eta}^s)$ as explained in the text:

```
forvalues s=1/$nclasses{
by $id:  egen double H_`s'=sum(h`s')
}
```

Before starting the loop that iterates the EM recursion until convergence, we need to specify a Stata *ml* command to perform the estimation of the grouped-data model defined in equation 5.30. and to do this, we use the Stata optimisation tool. In what follows we use four covariates (the *risk factors*) to model the probability of class membership. Finally bear in mind that the vector of parameters for one latent class is set to zero for identification:

```
*****************************************************
***compute the ml command for the grouped-data model***
*****************************************************
capture program drop logit_lf
program logit_lf args lnf xb za sa
qui replace `lnf'=[H1*ln(1/(1+exp(`xb')+exp(`za')+exp(`sa')))+
  H2*ln(exp(`xb')/(1+exp(`xb')+exp(`za')+exp(`sa')))+
  H3*ln(exp(`za')/(1+exp(`xb')+exp(`za')+exp(`sa')))+
  H4*ln(exp(`sa')/(1+exp(`xb')+exp(`za')+exp(`sa')))] if didep==1
qui replace `lnf'=0 if didep==0
end
```

We now present the loop that repeates the steps above until convergence:

```
****************
***Start loop***
****************
set more off
local i=1 while `i'<= $niter{
quietly{
drop l_*
**Estimate again the C CLOGIT models (one for each class) using the conditional (posterior)
probabilities as weights (as in eq.  5.29).  Then recompute the probability of each
alternative (the variables l_1,l_2,..,l_C) using the updated parameters:
set more off
forvalues s=1/$nclasses{
clogit $depvar $varlist [iw=H_`s'], group($idt)
predict double l_`s'
}
```

```
**Now call the ml_model defined before and maximise the grouped-data log-likelihood
defined in eq.  5.31.  Bear in mind that we introduce four covariates (defined as
f1, f2, f3 and f4):
ml model lf logit_lf (f1 f2 f3 f4) (f1 f2 f3 f4) (f1 f2 f3 f4)
ml max
**update the unconditional probabilities of class membership according to eq.  5.32:
replace prob1=1/(1+exp(_b[eq1:_cons]+_b[eq1:f1]*f1+_b[eq1:f2]*f2+
  +_b[eq1:f3]*f3+_b[eq1:f4]*f4)+exp(_b[eq2:_cons]+_b[eq2:f1]*f1
  +_b[eq2:f2]*f2+_b[eq2:f3]*f3+_b[eq2:f4]*f4)+exp(_b[eq3:_cons]+
  _b[eq3:f1]*f1+_b[eq3:f2]*f2+_b[eq3:f3]*f3+_b[eq3:f4]*f4))
replace prob2=exp(_b[eq1:_cons]+_b[eq1:f1]*f1+_b[eq1:f2]*f2+
  _b[eq1:f3]*f3+_b[eq1:f4]*f4)*prob1
replace prob3=exp(_b[eq2:_cons]+_b[eq2:f1]*f1+_b[eq2:f2]*f2+
  _b[eq2:f3]*f3+_b[eq2:f4]*f4)*prob1
replace prob4=exp(_b[eq3:_cons]+_b[eq3:f1]*f1+_b[eq3:f2]*f2+
  _b[eq3:f3]*f3+_b[eq3:f4]*f4)*prob1
**Now simply update the variables constructed before the loop in order to update
the conditional probabilities used as weights in the previous two sets of maximisations:
capture drop kbbb*
forvalues s=1/$nclasses{
replace kbb‘s’=l_‘s’*$depvar
qui recode kbb‘s’ 0=.
by $id:  egen double kbbb‘s’=prod(kbb‘s’)
by $id:  replace kbbb‘s’=.  if _n!=_N
}
replace den=prob1*kbbb1
forvalues s=2/$nclasses{ replace den=den+prob‘s’*kbbb‘s’
}
***update the weights using the conditional probabilities:
forvalues s=1/$nclasses{
replace h‘s’=(prob‘s’*kbbb‘s’)/den
recode h‘s’ .=0
}
drop H_*
forvalues s=1/$nclasses{
by $id:  egen double H_‘s’=sum(h‘s’)
}
```

Finally, show the value of the maximised log-likelihood (which is contained in the variable *sumll* defined below) and restart the loop:

```
capture drop sumll
egen sumll=sum(ln(den))
sum sumll global z=r(mean)
local i=‘i’ +1
}
display as green "Iteration " ‘i’ ":  log likelihood = " as yellow $z
}
```

Importantly, since EM algorithms do not compute Standard Errors, these can be obtained by applying the bootstrap command in Stata.

Part III.

Conclusions

6. Final Considerations

The aim of this thesis has been to present a behavioural microsimulation model for the Italian population. As we has seen in many occasions along the various chapters, behavioural models are a very useful tool for the analysis of tax-benefit reforms. Indeed, they represent a more general framework with respect to traditional static microsimulation models since they allow for individual behavioural reactions to changes in the economic environment that are induced by a given public policy.

The dissertation has explored a discrete approach for the estimation of the labour supply behaviour, which allows incorporating easily in the analysis any kind of nonconvexities in the budget sets and the non-participation choice. The model we have proposed includes several innovations with respect to other Italian studies on labour supply. In particular, we considered simultaneously endogenous childcare expenditures, unobserved preference heterogeneity, endogenous fixed costs of working and a systematic way to allow for the endogeneity of wages and errors in the prediction of wages for non-workers. Importantly, the estimated preference parameters define a utility function that is coherent with the economic theory, with an estimated marginal utility of consumption increasing at a decreasing rate for 99% of households in our sample. Moreover, women preferences for work are decreasing with the number of dependent children, in particular when the youngest child is under six years old. Furthermore, preferences for work decrease for either spouses when they are from southern Italy; increase with own age at a decreasing rate and are higher for people with high levels of education. The estimated average labour supply elasticities are in line with the empirical literature ranging between 0.62 - 0.89 for married women and between 0.15 and 0.28 for married men. Average own elasticities are higher in southern Italy, in particular for households with low household income, and they are lower for couples with young children.

The application of this model to the evaluation of the last three main reforms of the Italian personal income tax has shown small changes both in the level of inequality and in labour supply, something that was expected given the small revenue losses produced by these reforms. Hence, attention was focused on the sign, more than on the magnitude of the behavioural changes.

As for the distributive effects, we found that the two centre-right modules, which had a total cost of about 13 billion euros, had very small distributive impacts overall, whilst the subsequent reform by the centre-left government produced a greater amount of redistribution, since it actually increased the tax burden for high incomes, with no revenue loss. From a distributive point of view, therefore, the difference between the two approaches is clear, although in all cases the changes in the Gini index have been quite modest.

Concerning the behavioural reactions, the adjustment in the tax structure with the most significant consequences in terms of labour supply incentives is the extension of the no-tax area in 2003. This reform produced an increase in female labour supply for low and middle deciles. As a consequence, the behavioural reactions to this reform increased real incomes at the bottom of the distribution, therefore further reducing the Gini index. Hence, the 2003 module produced both a reduction in inequality and an increase in labour supply, showing complementarity between equity and efficiency. The 2005 module, on the other hand, increased inequality but had a (smaller) additional positive impact on labour supply, that slightly reduced the tendency for inequality to rise. Interestingly, the behavioural contribution that in part counterbalanced the rise in inequality comes from the lower deciles since the behavioural impact of this reform on the top deciles is absolutely negligible. The 2007 reform, finally, has had a clear equity effect, further reducing inequality, but with a reduction in efficiency, particularly among low-income women. This could derive from the expansion of cash transfers, that are decreasing with family income and therefore produce both an income and a substitution effect on the choice between leisure and consumption, in particular for women in couples with children. The 2007 reform actually presents the traditional trade-off between efficiency and equality since it concentrates more public funds towards low and middle-income households, that have a relatively elastic labour supply.

Importantly, comparing the final structure of the 2007 income tax with that of the baseline year (2002), several steps forward emerge. Overall work incentives have improved for both women and men in couples, without reducing the redistributive effect of the personal income tax. From this analysis we derived important policy recommendations, suggesting the policy makers to reduce marginal tax rates on low incomes, providing positive effects in terms of both income distribution and labour supply, while child benefits should not decrease too rapidly with the level of family income.

In the last chapter of this essay we focus not on a new application of our model, but on a practical issue related to this kind of behavioural modelling. Indeed, we have explored in deep detail the role of unobserved preference heterogeneity in discrete labour supply models and found that the post-estimation results as elasticities, inequality index and behavioural responses to tax changes across the population might differ significantly depending the the way researchers account for unobserved preference variability. Since standard gradient-based optimisation algorithm fails to achieve convergence when unobserved heterogeneity in considered in a more general and comprehensive way, we have developed a new optimisation method based on an EM algorithm that allows for the nonparametric estimation of mixing distribution in discrete choice models of labour supply.

Through our algorithm, we compared post-estimation results for different specifications of unobserved heterogeneity and found that results do change significantly with respect to the base model without unobserved heterogeneity, only when a large set of parameters is allowed to vary randomly and nonparametrically in the population. Moreover, we have simulated the distributive and efficiency effects of the introduction of a UK-style working-tax credit into the Italian tax-benefit system and found that also policy recommendations related to given tax-reforms may change significantly according to the specification of the unobserved taste variability in the model.

Our future research will be based on the extension of the behavioural microsimulation model. In particular, the estimation of discrete labour supply models with endogenous bargaining for married couples will play an important role in the future research. What is more, the inclusion of dynamic elements and labour market rationing are crucial extensions that have to be introduced in order to relax some of the assumptions that underline our model.

Bibliography

Aaberge, R., Colombino, U., Strøm, S., 2000. Labor supply responses and welfare effects from replacing current tax rules by a flat tax: Empirical evidence from Italy, Norway and Sweden. Journal of Population Economics 13 (4), 595–621.

Aaberge, R., Colombino, U., Strom, S., Wennemo, T., 1998. Evaluating alternative tax reforms in Italy with a model of joint labor supply of married couples. Structural Change and Economic Dynamics 9 (4), 415–433.

Baldini, M., 2004. MAPP02: un modello di analisi delle politiche pubbliche. Dipartimento di economia politica, Modena, Materiale di discussione 331.

Baldini, M., Bosi, P., 2002. La riforma dell'imposta sul reddito: aspetti di equità e di efficienza. Politica economica (3), 303–340.

Baldini, M., Bosi, P., Toso, S., 2002. Targeting welfare in Italy: old problems and perspectives on reform. Fiscal Studies 23 (1), 51–75.

Baldini, M., 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., 2008. Gli effetti distributivi dei trasferimenti in kind: il caso dei servizi educativi e sanitari. A. Brandolini e C. Saraceno (a cura di), Povertà e benessere. Una geografia delle disuguaglianze in Italia, 401–423.

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.

Berliri, C., Parisi, V., 2006. The Distributive and Labor Supply Impact of the Minimum Insertion Income: The Case of Italy. International Advances in Economic Research 12 (2), 213–227.

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., Chiappori, P., Magnac, T., Meghir, C., 2008. Collective Labour Supply: Heterogeneity and Non-Participation. The Review of Economic Studies 74 (2), 417–445.

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. Handbooks in Economics 5 (3 PART A), 1559–1696.

Brandolini, A., 1999. The distribution of personal income in post-war Italy: source description, data quality, and the time pattern of income inequality. Giornale degli Economisti 58 (2), 183–239.

Breunig, R., Cobb-Clark, D., Gong, X., 2008. Improving the Modelling of Couples Labour Supply. The Economic Record 84 (267), 466–485.

Brewer, M., Duncan, A., Shephard, A., Suarez, M., 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.

Cappellari, L., Jenkins, S., 2003. Multivariate probit regression using simulated maximum likelihood. The Stata Journal 3 (3), 278–294.

Chiappori, P., 1992. Collective labor supply and welfare. Journal of Political Economy 100 (3), 437.

Chiappori, P., Ekeland, I., 2006. The micro economics of group behavior: general characterization. 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., Duncan, A., 2002. Behavioural microsimulation with labour supply responses. Journal of Economic Surveys 16, 1–39.

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.

De Vincenti, C., Paladini, R., Pollastri, C., 2005. For a welfare-oriented taxation reform in Italy. Giornale degli Economisti e Annali di Economia 64 (2/3), 189.

Del Boca, D., Locatelli, M., Vuri, D., 2005. Child-Care Choices by Working Mothers: The Case of Italy. Review of Economics of the Household 3 (4), 453–477.

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.

Duncan, A., Stark, G., 2000. A recursive algorithm to generate piecewise linear budget contraints. IFS Working Papers.

Gastaldi, F., Liberati, P., 2005. The Personal Income Tax in Italy: Why Does It Change? Giornale degli Economisti 64 (2-3), 159–188.

Gould, W., Pitblado, J., Sribney, W., 2006. Maximum Likelihood Estimation with Stata. Stata Press books.

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

Greene, W., 2007. Econometric Analysis (7 th).

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., Singer, B., 1984. A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52 (2), 271–320.

Heim, B., Meyer, B., 2004. Work costs and nonconvex preferences in the estimation of labor supply models. Journal of Public Economics 88 (11), 2323–2338.

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

Labeaga, J., Oliver, X., Spadaro, A., 2008. Discrete choice models of labour supply, behavioural microsimulation and the Spanish tax reforms. Journal of Economic Inequality 6 (3), 247–273.

Mancini, A., 2007. Labor Supply Responses of Italian Women to Minimum Income Policies. CHILD Working Papers.

Manning, A., Petrongolo, B., 2008. The Part-Time Pay Penalty for Women in Britain. The Economic Journal 118 (526), F28–F51.

Mazzaferro, C., Morciano, M., 2008. CAPP_DYN: A Dynamic Microsimulation Model for the Italian Social Security System. Center for the Analysis of Public Policies (CAPP).

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 Applied 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 de Ven, J., Weale, M., 2009. A Structural Dynamic Micro-Simulation Model for Policy Analysis: Application to Pension Reform, Income Tax Changes and Rising Life Expectancy. NIESR Discussion Papers.

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.