CAPP

A behavioral microsimulation model
with discrete labour supply for Italian couples

Daniele Pacifico

A behavioral microsimulation model with discrete labour supply for Italian couples

Daniele Pacifico*

Abstract

The aim of this paper is to introduce labour supply behaviour in an arithmetic microsimulation model so as to take into account changes in labour supply when a new policy is evaluated. I 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 instead of labour supply functions. The main advantage of the discrete approach is the possibility of dealing easily with non-convex budget sets and joint labour supply. This let the discrete approach relatively suitable for policy evaluation purposes. I 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. Several innovative elements are taken into account with respect previous Italian studies. In particular, I allow for errors in the predicted wage for non-workers, unobserved heterogeneity in preferences, unobserved monetary fixed costs of working and child-care demand. The model is fully parametric and the Simulated Maximum Likelihood approach is used to approximate multidimensional integrals. An overview of the STATA routine for the maximum likelihood estimation is also presented. The elasticities of labour supply for married men and women are computed and discussed.

*PhD candidate at the University of Bologna. Research fellow of CAPP, University of Modena and Reggio Emilia. Member of CRENoS, Cagliari. Contact: daniele.pacifico@unibo.it Web: www.danielepacifico.com

Introduction

The aim of this paper is to present a microeconometric labour supply model for policies evaluation purposes[1]. There exist several other models that simulate the Italian labour supply and this is another contribute in this direction. Differently from most of the labour supply literature of the past, I use a discrete choice model of labour supply. As Blundell and MaCurdy (1999) point out, 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. In fact, the discrete approach allows for some important extensions and relaxes strong assumptions commonly used in other estimation techniques.

Traditionally, structural labour supply models have been estimated assuming a choice set defined on any positive real number of worked hours. This is what Van Soest (1995) defines as the *continuous approach*. In a static framework, this means that the agent chooses the best combination of consumption and leisure so as to maximize her utility function given a time and a budget constraint. Notice that 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 of available time. It is worth to notice how this optimization program could be extremely cumbersome to estimate. The literature has developed two different approaches in continuous microsimulation. 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 priori* 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 has to be estimated with the subsequent difficulties during the integration procedure. Another common method in continuous microsimulation is to work directly with preferences with supply function derived directly from either a direct or an indirect utility function. Here the main problem is the tax schedule in the budget constraint that 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 constraints for each possible level of labour supply. In continuous models 1 or 5 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[2]. Another complicated issue is the presence of a tax-benefit system that may give rise to highly non-linear and non-convex budget sets for most of the population of interest. It follows that feasible estimations require the linearization of the budget constraint around

---

[1]The author is grateful to Monica Costa Dias and Mike Brewer for their advices and comments. Thanks to Massimo Baldini for providing me the algorithms of his tax-benefit simulator.

[2]Duncand and Stark (2000) have developed an algorith in GAUSS that is able to recover budget sets more efficiently and accurately.

the observed hours level or the construction of search-algorithms that compare the maximum utility on each linear segment of a piecewise-linear budget constraint[3]. Moreover, considerable problems arises because of the simultaneity between net wages and hours worked due to the tax-schedule with the subsequent necessity of finding appropriate instruments so as to ensure identification. Finally, other difficulties arise when the model try to allow for important extensions like unobserved preference heterogeneity or joint labour supply. As Creedy and Duncan (2002) point out, these criticisms make the continuous approach seldom used nowadays. Instead, the discrete approach avoids most of these problems and has several other advantages. It is based on the assumption of utility maximizing 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 hour choice 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 net-incomes, the tax schedule does 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 made 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 of hours. But the discrete approach has other important advantages. In particular, it allows for important extensions that are difficult to consider in the standard, 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 and joint labour supply could be incorporated in the model in a convenient way. The main drawbacks of the discrete approach are the rounding error produced when the choice set is discretized as well as the incomplete use of available information[4]. 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 public policy. Earlier international works that explore this method are those from Van Soest (1995), Keane and Moffitt (1998) and Blundell, Dancan, McCrae and Meghir (1999). 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

---

[3]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 searcing for the optimal level of labour supply, allowing for nonlinearities and nonconvexities. See Creedy and Duncan (2002)for references.

[4]The wider are the hours categories used to discretise the choise set the bigger is the rounding error, see Van Soest (1995).

et al. (2006) who extend the paper of Blundell et al. (1999) to study the impact of the WFTC reforms in the UK, Breunig et al. (2005) who estimate the wage equation and the structural labour supply model simultaneously allowing for correlation between the random terms, Haan (2004) who studies the German case comparing the performance of a random coefficients specification with respect to the performance of a more simple conditional logit model, Labeaga et al.(2007) who study the impact of the Spanish tax reform on efficiency and social welfare. A very active centre that is specialized 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, Colombino and Stroem (1999) 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(2008), developed a model that is closely related to the one discussed here in order to study labour supply responses to minimum income policies. Del Boca and Vuri (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. Differently from the other models that have been developed for the Italian case, the present one considers many innovative features simultaneously. Indeed, I estimate jointly the labour supply behaviour of married men and women allowing for errors in predicted wages, work-related monetary costs, child-care costs and unobserved random preferences. Work-related monetary costs are important since they eliminate the reservation wage condition in estimation. Moreover, depending on how these costs are specified, they may help to relax the assumption of fixed wage rates. Differently from previous Italian works, I also take into account child-care costs and unobserved heterogeneity in preferences. Both these aspects are relevant. From a practical point of view, child-care costs have a role similar to those of other fixed costs of working since they may help to eliminate the reservation wage condition 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 I take into account the endogeneity related to estimated wages for non-workers. Indeed, I integrate out wages by drawing randomly from their estimated distribution and weighting the likelihood when wages are not observed in the data.

This paper is structured as follows. Section 1 presents the general framework for the structural microeconometric model. Section 2 presents the extensions to the basic model. Section 3 explains the data used for the empirical analysis.

---

[5] As it will be clarified later, the direct utility function is assumed to follow a type 1 extreme value distribution that underlines the typical IIA (independence from irrrelevant alternatives) assumption.

4

Section 4 describes the estimation procedure. Section 5 contains results from first stage regressions and section 6 discusses the estimates of the structural model. The appendix contains an overview of the Stata algorithm coded for the estimation of the structural model.

1. The basic econometric model

In this section I develop the econometric framework for the empirical analysis. I focus only on married/*de facto* couples and do not consider singles[6]. It follows that the couple has to be considered as the decision maker[7]. The couple chooses a particular combination of hours of work for both the spouses in order to maximize a joint utility function defined over the household net-income and the hours of work of both the spouses. As standard in the literature, I 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 maximizing and that the utility is not deterministic, it is possible to recover the probability of a particular choice. This is the base for the computation of the likelihood function.

To be formal, let $\mathbf{H} = [hf; hm]$ be a vector of worked hours, hf for women in couples and hm for men in couples. Let Y_{H^j} be the net household income and X be a vector of individual characteristics. Then the household utility for a particular choice $H = H^j$ can be defined as:

$$U_{H^j} = U(Y_{H^j},\, H^j,\, X) + \xi_{H^j} \tag{1}$$

Where ξ_H is a choice-specific stochastic component which is assumed 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 optimization error. The net-household income Y_{H^j} when the vector $H = H^j$ is chosen is defined as:

$$Y_{H^j} = \bar{w}_f \cdot hf^j + \bar{w}_m \cdot hm^j + Nly + TB(\bar{w}_f^j; \bar{w}_m^j; H^j; Nly; X) \tag{2}$$

Where $\bar{w}_f$ and $\bar{w}_m$ are the (fixed) hourly gross wages from employment for women and men respectively; Nly is the non-labour income and the function

---

[6]The model for single is under development and will be available soon. The strategy is almost the same as it will be clarified later.

[7]In this paper use a unitary model of labour supply. A recent literature has shown that the unitary assumption is often rejected. Nevertheless, a collective model is far away to be a practicable model. Moreover, a collective model has to be simplified in other directions and is based on discountable assumtions for the identification of the bargaining parameter. However, it is challenging to develop and make it practicable a collective model of labor supply.

[8]Some studies have found a part time pay penalty, see Manning and Petrongolo (2004). However there are several ways that relax this assumption that will be discussed in the next sections.

$TB(\bar{w}_f; \bar{w}_m; H; Nly; X)$ represents the tax-benefit system which depends on the gross wage rates, hours of work, household non-labour income and individual characteristics. It is worth to notice that this function can be highly non-linear for most of the population of interest. Following Keane and Moffitt (1998) and Blundell et al. (1999), the utility above is defined as a second order polynomial with interaction between the wife and the husband terms:

$$U(y_{H^j};\ H^j; X) = \begin{aligned} &\alpha_1 y_{H^j}^2 + \alpha_2 hf_j^2 + \alpha_3 hm_j^2 + \alpha_4 y_{H^j} hf_j + \alpha_5 y_{H^j} hm_j + \\ &+\alpha_6 hf_j hm_j + \beta_1 y_{H^j} + \beta_2 hf^j + \beta_3 hm^j + \xi_{H^j} \end{aligned}$$

(3)

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

$$\beta_j = \sum_{i=1}^{K_j} \beta_{ji} x_{ji} + \nu_j \qquad j \in \{1, 2, 3\}$$

with ν_j terms being unobserved household preferences that are assumed to be independent 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 the model.

Under the assumption that the couple maximizes her utility over a discrete set of alternatives and that the error term in the utility function follows a type one extreme value distribution, the probability of choosing a particular vector $H^j = [h_f^j;\ h_m^j]$ is given by[10]:

$$\begin{aligned} Prob(H = H^j | X, \boldsymbol{\nu}) &= Pr[U(Y_{H^j}, H^j, X, \nu) > U(Y_{H^s}, H^s, X, \nu), \forall\, s \neq j] \\ &= \frac{exp(U(Y_{H^j},\ H^j,\ X,\ \nu))}{\sum_{k=1}^{K} exp(U(Y_{H^k},\ H^k,\ X,\ \nu))} \end{aligned}$$

(4)

Where $\boldsymbol{\nu} = (\nu_1, \nu_2, \nu_3)$.

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:

---

[9]IIA is the acronym of Independence form Irrelevant Alternatives (McFadden). The main problem with this assumption is that the probability of choosing alternative j is always the same as the probability of choosing alternative i, for any i and j. This assumption is particularly restrictive in the labour supply framework. Consider a choice set initially defined by just two alternatives: working full time and not working. The IIA assumption implies that introducing another alternative, say a part time job, does not change the relative odds between the two initial alternatives.

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

$$\mathcal{L}_i = \int_{\boldsymbol{\nu}} \prod_{j=1}^{K} d_{ij} \left(\frac{exp(U(Y_{H^j}; H^j; X, \nu))}{\displaystyle\sum_{k=1}^{K} exp(U(Y_{H^k}; H^k; X, \nu))} \right) \phi(\boldsymbol{\nu}) d(\boldsymbol{\nu}) \qquad (5)$$

Where d_{ij} is a dummy variable equal to one for the observed choice and zero otherwise. Up to this point, I 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]. I assume that wage for agent i is generated by the following standard Heckman selection process:

$$log(w_i) = X_{1i}\beta + \epsilon_i \qquad (6)$$

$$w_i \text{ is observed} \iff U_i^* = utility \ of \ work > 0$$

$$U_i^* = Z_i\alpha + \nu_i \qquad (7)$$

$$\begin{pmatrix} \epsilon_i \\ \nu_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_\epsilon^2 & \rho \\ \rho & 1 \end{pmatrix} \right) \qquad (8)$$

Where $Z_i = [X_{1i} X_{2i}]'$ 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:

$$\mathcal{L}_i = \int_\epsilon \int_{\boldsymbol{\nu}} \prod_{j=1}^{K} d_{ij} \left(\frac{exp(U(Y_{H^j}; H^j; X))}{\displaystyle\sum_{k=1}^{K} exp(U(Y_{H^k}; H^j; X))} \right) \phi(\epsilon)\phi(\boldsymbol{\nu}) d(\epsilon) d(\boldsymbol{\nu}) \qquad (9)$$

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

---

[11] See (Keane and Moffitt, 1998).

2. 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 and improve the fitting[12]. There are several ways to account for these problems. For example, Aaberge et al. (1999) 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 requiring 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 an 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 I 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, I consider the characteristics of different discrete points by estimating the monetary cost of working for three groups of discrete points. In particular, I 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 I 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

---

[12]Indeed, it has been found that the basic model sistematically overpredict 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, Colombino, and Stroem (1999) for details.

[14]This approach has been follrwen by several authors, see Mancini (2008), Haan (2004) and Van Soest (1995)

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

point that actually drives also the specification in Aaberge et al. (1999). 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]:

$$FC(hf_j, Z) = Z_1\gamma_1 \cdot 1\{hf_j > 0\} + Z_2\gamma_2 \cdot 1\{hf_j > 30\} \qquad (10)$$

Where Z_1 and Z_2 are vectors of individuals characteristics, γ_1 and γ_2 are vectors of parameters estimated jointly with the other structural parameters and $1\{-\}$ are binary indicators that take value one whenever the argument inside the brakets is true.

To take into account child-care costs I adopt a different strategy. As pointed out in Del Boca and Vuri (2005), Italy has an objective lack of data on child-care usage and child-care costs. In order to overcome this lack of data, I recovered information on child-care costs from another dataset. In particular, I computed the hourly price of child-care for different groups of households and for each group I 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, I 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. I 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. I also 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 dataset. 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 likelihoodto account for not-observed quality. Formally, we can define a child-care cost function as:

$$CC(hf_j, p_c, X) = E[h_{cc}|X, hf_j] \cdot p_c \qquad (11)$$

Where p_c is a particular price for an hour of child-care and $E[h_{cc}|X, h_f^j]$ 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 in the model as a once off cost directly subtracted from income at any possible choice of hours. Defining a total cost function as:

$$TC = CC(hf_j, p_c, X) + FC(hf_j, Z) \qquad (12)$$

---

[16]See Heim and Meier (2004).
[17]As in Blundell et al. (1999) or Brewer et al. (2006), I assume positive costs of working only for female.

the utility function changes as follows:

$$U_{H^j} = [U(Y_{H^j} - TC; H^j; X)] + \epsilon_{H^j} \tag{13}$$

and the likelihood for observation i becomes:

$$\mathcal{L}_i = \sum_{s=1}^{5} P(p_c^s|X) \int_{\boldsymbol{u}} \prod_{j=1}^{K} d_{ij} \left(\frac{exp(U(Y_{H^j} - TC; H^j; X, \nu))}{\sum_{k=1}^{K} exp(U(Y_{H^k} - TC); H^k; X, \nu))} \right) \phi(\boldsymbol{u})d(\boldsymbol{u})$$
$$\tag{14}$$

Where $\boldsymbol{u} = (\epsilon, \nu_1, \nu_2, \nu_3)$ and d_j is a dummy that pick up the observed choice. The likelihood above is difficult to estimate since it requires the computation of a four dimensional integral. Following Train (2003), I apply simulation methods to approximate these integrals. In particular, I use Halton sequences instead of traditional random draws from the densities. Given that Halton sequences ensure a more complete coverage of the integration support, less draws are needed to reach consistent estimates other things being equal. The simulated log-likelihood is:

$$L = \sum_{n=1}^{N} ln \frac{1}{R} \sum_{r=1}^{R} \sum_{s=1}^{5} P(p_c^s|X) \prod_{j=1}^{K} d_{ij} \left(\frac{exp(U(Y_{H^j} - TC^s; H^j; X, \nu^r))}{\sum_{k=1}^{K} exp(U(Y_{H^k} - TC^s); H^k; X, \nu^r))} \right)$$
$$\tag{15}$$

The routine I coded to compute this likelihood is discussed in the appendix.

3. The 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 I use the cross sectional survey for the year 2002. The dataset 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, I focus only on a selected sub-sample of the whole population. In particular, as standard in the literature on labour supply, I do not consider couples with spouses who are

---

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

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 behavior in the labour market that is not characterized 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 I use a modified version of the arithmetic tax-benefit microsimulation model called $MAPP$02. 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, I make use of another dataset 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, 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 on child-care expenditure, hours of child-care and hours of work. Unfortunately, the information on child-care expenditures is registered only for children aged less than 6 so that I am 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 free education for older children. The two datasets are relatively similar and both are representative of the same population.

4. Estimation procedure

In this section I 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 section. The first step is the definition of the relevant sub-sample and the re-arrangement of the information contained in the two datasets so as to gather all the information needed. 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. The gross hourly wage for the unemployed is estimated making use of the wage model 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 Heckman model it is possible to predict hourly gross wages for non-workers. Following the wage

---

[19]See Baldini (2002).

model outlined above, predicted wages for non-workers could be computed as:

$$E[ln(w_i)|ln(w_i) \text{ is not observed}; X] = X_i\hat{\beta} - \hat{\sigma}_\epsilon\hat{\rho}\lambda(Z_i\hat{\alpha}) \tag{16}$$

Nevertheless, using just the predicted wage for non-workers would lead to inconsistent estimates as long as wages are endogenous and predicted with errors. Here, I 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]:

$$\epsilon_i \mid ln(w_i) \text{ not observed} \sim N\left(-\hat{\sigma}_\epsilon\hat{\rho}\lambda(Z_i\hat{\alpha}) \; ; \; \hat{\sigma}_\epsilon^2\left[1 - \hat{\rho}^2\psi(Z_i\hat{\alpha})\right]\right)$$

Where: $\psi(Z_i\alpha) = \lambda(Z_i\alpha)\left(\lambda(Z_i\alpha) + Z_i\alpha\right)$ and $\lambda(Z_i\alpha) = \frac{\phi(Z_i\alpha)}{1-\Phi(Z_i\alpha)}$. I drew 50 pseudo-random numbers for each observation from this latter distribution. Then I added 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-workers i is:

$$ln(w_i) = X_i\hat{\beta} + \epsilon_i^r \tag{17}$$

Finally, notice that:

$$exp(E[ln(x)]) \neq E[exp(ln(x))] = E[x]$$

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

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

which in our case means:

$$E[w_i|w_i \text{ not obs}; X] = exp(X_i\hat{\beta} + \epsilon_i^r)exp(\frac{\hat{\sigma}_\epsilon^2}{2}\left[1 - \hat{\rho}^2\psi(Z_i\hat{\alpha})\right]) \tag{18}$$

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 discretized 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].

---

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

[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%)
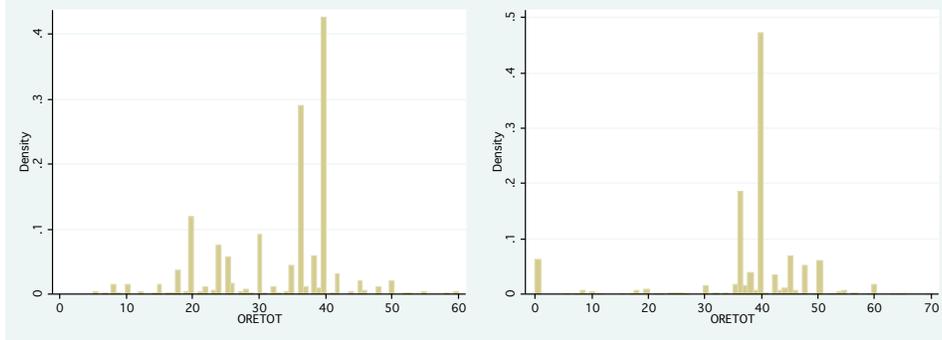
According to these distributions, women in a couple are restricted to choose from the following discrete set: x={0, 10, 20, 30, 40}. These points correspond to the following intervals: 0-5, 6-15, 16-25, 26-36, >36. For married men I selected the following discrete set: y={0, 40, 50} that corresponds to the intervals 0-10, 11-42, >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 lead 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 n possible choices of hours and *50* draws each non-worker has *n50* 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 dataset. In this latter dataset I first drop the households without children younger than 6 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. Using only this sub-sample, I define 8 groups for child-care expenditure according to the presence of children aged less than 3, geographic area (north, south) and mother's education (low or high). For each group I computed the distribution of hourly expenditure in formal child-care for those households observed demanding formal child-care. Then I approximate this distribution using 4 mass points. Next, I computed sub-groups controlling for the mother's working status, which implies that 16 groups are now defined. For each of them I computed 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. In practice, this means that for a mother in a particular group

13

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 given by:

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

(19)

Where $Pr(H^j|X, \boldsymbol{u}, p_c)$ 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., 1999)I 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:

$$h_{cc} = \beta_0 + \beta_1 h f_j$$

(20)

Whit this information it is possible to estimate the structural model. The estimation algorithm is implemented in STATA. An overview of the Stata routine I built up for this work is in the appendix

5. 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 the south of 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 to notice 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 to point out that the correlation coefficient between the two error terms - rho - is statistically different from zero and positive in both the equations.

Table 1. Wage equations

	Female		Male	
	coef	z	coef	z
Log gross hourly wage:				
$educ2^{\S}$	0.299	4.390	0.197	5.790
$educ3^{\S}$	0.681	7.990	0.422	12.140
$educ4^{\S}$	1.087	10.190	0.825	17.510
$Age^{\dagger}$	-0.004	-0.270	0.038	2.840
$Age\,squared^{\dagger}$	0.021	1.020	-0.029	-1.900
$Area1:Northern^{\S}$	0.244	4.080	0.262	9.600
$Area2:Middle^{\S}$	0.158	2.800	0.157	4.930
$Home\,owner^{\S}$	0.125	3.410	0.111	4.900
_const	1.246	3.440	0.658	2.240
Selection equation:				
$Net\,income\,at\,0\,hours$	-0.109	-2.590	-0.218	-4.070
$educ2^{\S}$	0.439	4.190	0.328	2.520
$educ3^{\S}$	1.150	10.830	0.523	3.690
$educ4^{\S}$	1.870	12.460	1.354	3.420
$Age^{\dagger}$	0.095	2.560	0.190	3.380
$Age\,squared^{\dagger}$	-0.001	-2.860	-0.002	-3.520
$Home\,owner^{\S}$	0.424	6.120	0.360	3.200
$Area1:Northern^{\S}$	0.948	12.990	0.864	5.970
$Area2:Middle^{\S}$	0.708	8.140	0.811	4.840
$Age\,youngest\,child:0/2^{\S}$	-0.436	-3.710	0.326	1.510
$Age\,youngest\,child:3/5^{\S}$	-0.377	-3.100	0.385	1.710
$Age\,youngest\,child:6/9^{\S}$	-0.330	-2.910	0.370	1.750
$Age\,youngest\,child:10/16^{\S}$	-0.104	-1.310	0.243	1.890
_const	-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. $\dagger$ denotes that the variable is measured in terms of deviation from its mean. $\S$ denotes discrete variables. Educ$\sharp$ 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 2. distribution of hourly child-care cost (euro)

Group:	qtile: 12.5	qtule: 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. The and that use formal childcare.

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. 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 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.

6. Results from the structural model

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 5. Structural model estimates for couples

	parameter	Coef.	z
α_1 :	constant	-0.084	-7.480
α_2 :	constant	-0.175	-3.330
α_3 :	constant	-0.384	-25.380
α_4 :	constant	0.559	3.030
α_5 :	constant	-0.160	-9.100
α_6 :	constant	0.033	1.720
β_1 :	constant	2.571	12.610
	Wife's age†	-0.039	-0.450
	Husband's age†	0.131	2.300
	Southern Italy§	0.220	1.910
	Wife's education (high)§	-0.247	-2.500
	Husband's education (high)§	-0.016	-0.340
	Number of children	-0.087	-1.160
	Youngest child 0-6§	-0.030	-0.190
	σ_1	0.157	2.432
β_2 :	Constant	2.112	6.430
	Wife's age†	0.713	3.610
	Wife's age squared†	-0.092	-4.060
	Southern Italy§	-0.189	-2.030
	Wife's education (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
β_3 :	Constant	1.386	13.260
	Husband's age†	0.544	2.200
	Husband's age squared†	-0.079	-2.800
	Southern Italy§	-0.248	-3.990
	Husband's education (high)§	0.011	0.210
	Number of children	0.065	1.530
	Youngest child 0-6§	0.055	0.620
	σ_3	0.024	0.168
FC_f^1	Constant	2.667	10.680
FC_f^2	Constant	1.161	9.800

Log-Likelihood: -3348.3188   Observations: 2002 couples

Note: model estimated by Simulated Maximum Likelihood using Halton sequences (50 draws). Annual household income divided by 1000; Women and men's worked hours divided by 10; Random terms divided by 10; α2 and α3 divided by 100; α4 divide by 1000. § denotes dummy variables and † denotes that variables are measured in terms of deviation from their means. σ coefficients are estimated standard deviations. FC1 represent fixed costs of working. FC2 represents additional fixed costs of working for full-time jobs.

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 both the 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 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. 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:

$$Pr(H^j|Y^p, X) = \frac{1}{R}\sum_{r=1}^{R}\sum_{s=1}^{5} P(p_c^s|X)\frac{exp(U(Y_{H^j}^p - TC^s; H^j; X, \nu^r))}{\sum_{k=1}^{K} exp(U(Y_{H^k}^p - TC^s); H^k; X, \nu^r))}$$

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

$$E[H_s|Y^p \, , \, X] = \sum_{k=1}^{K^s} Pr(H^k|Y^p \, , X) \cdot H_s^k$$

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

$$\varepsilon_s = \frac{E[H_s|Y^{after} \, , \, X] - E[H_s|Y^{before} \, , \, X]}{E[H_s|Y^{before} \, , \, X]} \cdot \frac{1}{0.01}$$

With *s=men, women.* The next table shows such elasticities. However, it is worth to notice that such elasticities have to be interpreted carefully. They are a useful summary measure of the labour supply behavior 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.

Table 7. Labour supply elasticities by individual characterisitcs.

	Husband's gross wage +1%	Wife's gross wage +1%
All married couples	0.151	0.87
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	.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

---

[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.

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

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.

Table 8. Average elasticities by 10 quantiles of equivalent income

Deciles of gross income	Husband's gross wage +1%		Wife's gross wage +1%	
	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

Conclusions

This essay has explored the performance of the discrete approach for the estimation of labour supply for married couple using Italian data. The discrete approach has several advantages with respect the continuous approach. In particular, it easily allows for highly non-linear budget sets and joint labour supply. Several innovations have been introduced with respect earlier Italian studies. In particular, I take into account unobserved heterogeneity in preferences, child-care demand, 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 observations and preferences for work decrease with the number of young children and increase with own 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 the southern of Italy and they are lower for couples with young children. Future work will extend the present labour supply model to singles. Finally, it would be challenging to develop a collective model of labour supply that takes into account bargaining between the spouses.

References

Aaberge, R., U. Colombino, and S. Stroem, "Labour supply in Italy: an empirical analysis of joint household decisions, with taxes and quantity constraints," *Journal of Applied Econometrics*, 1999, *14* (4), 403–422.

Baldini, M. and P. Bosi, "Flat Rate Tax, Dividendo sociale e riforma dei programmi di spesa di assistenza," *Rivista internazionale di scienze sociali*, 2002.

_ , _ , **and D. Pacifico**, " Gli effetti distributivi dei trasferimenti In Kind: il caso dei servizi educativi e sanitari," *Poverta e Benessere, una geografia delle diseguaglianze in Italia*, 2007, pp. 401–422.

Blundell, R., A. Duncan, J. McCrae, and C. Meghir, "Evaluating In-Work Benefit Reform: The Working Families Tax Credit in the UK," *Institute for Poverty Research Conference, North Western University, November*, 1999.

_ **and T. MaCurdy**, "Labor Supply: A Review of Alternative Approaches," *HANDBOOKS IN ECONOMICS*, 1999, *5* (3 PART A), 1559–1696.

Boca, D. Del and D. Vuri, "Labor Supply and Child Care Costs: The Effect of Rationing," 2005.

Brandolini, A., *The Distribution of Personal Income in Post-war Italy: Source Description, Data Quality, and the Time Pattern of Income Inequality*, Banca d'Italia, 1999.

Breunig, Cobb-Clark, and Gong, "Improving the Modeling of Couples' Labour Supply," CEPR Discussion Papers 499, Centre for Economic Policy Research, Research School of Social Sciences, Australian National University September 2005.

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

Cappellari, L. and S.P. Jenkins, "Calculation of Multivariate Normal Probabilities By Simulation, With Applications to Maximum Simulated Likelihood Estimation."

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

_ **and G. Kalb**, "Discrete Hours Labour Supply Modelling: Specification, Estimation and Simulation," *Journal of Economic Surveys*, 2005, *19* (5), 697–734.

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

Green, W.H., *Econometric Analysis (7 th)*, Upper Saddle River, NJ: Prentice Hall, 2007.

Haan, P., *Discrete Choice Labor Supply: Conditional Logit Vs. Random Coefficient Models*, DIW, 2004.

Keane, M. and R. Moffitt, "A Structural Model of Multiple Welfare Program Participation and Labor Supply," *International Economic Review*, 1998, *39* (3), 553–589.

Labeaga, JM., X. Oliver, and A. Spadaro, "Discrete Choice Models of Labour Supply, Behavioural Microsimulation and the Spanish Tax Reforms," *Journal of Economic Inequality.*

Mancini, A.L., "Labor Supply Responses of Italian Women to Minimum Income Policies," *Working Paper 94, ECINEQ, Society for the Study of Economic Inequality.*, 2008.

Manning, A. and B. Petrongolo, *The Part-time Pay Penalty*, Department of Trade and Industry, 2004.

McFadden, D., *Conditional Logit Analysis of Qualitative Choice Behavior*, Institute of Urban and Regional Development, University of California, 1973.

Soest, A. Van, "Structural Models of Family Labor Supply: A Discrete Choice Approach," *JOURNAL OF HUMAN RESOURCES*, 1995, *30* (1), 63–88.

Train, K., *Discrete Choice Methods with Simulation*, Cambridge University Press, 2003.

Appendix

Overview of the STATA routine

This section explains the routine I wrote for the optimization algorithm. Stata has a really powerful optimization routine which is able to use, even within the same searching process, different optimizations algorithms such as Newton-Raphson, BHHH, DFP etc. There are three possible methods to estimate a self-written likelihood. The first one, so-called method d0, requires the analyst to provide just the likelihood function. The second, 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 negative 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. I chose method d0 for my 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, I follow Train (2003) and use Simulated Maximum Likelihood with Halton sequences. The Stata command *mdraws* by Cappellari and Jenkins(2007) helps to generate the Halton Sequences from which it is easy to get the correspondent draws from the multivariate normal distribution. I 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 I defined in the Stata routine is:

$$U(y_{H^j};\ H^j; X) = \begin{aligned} & \alpha_1 y_{H^j}^2 + \alpha_2 h_f^{j2} + \alpha_3 h_f^{j2} + \alpha_4 y_{H^j} h_f^j + \alpha_5 y_{H^j} h_m^j + \alpha_6 h_f^j h_m^j \\ & + (\boldsymbol{\beta_1}\mathbf{x_1'} + v_1)y_{H^j} + (\boldsymbol{\beta_2}\mathbf{x_2'} + v_2)h_f^j + (\boldsymbol{\beta_3}\mathbf{x_3'} + v_3)h_m^j + \xi_{H^j} \end{aligned} \tag{21}$$

where:

$$\nu \sim N\left(\begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}, \begin{pmatrix} \sigma_{1,1}^2 & 0 & 0 \\ 0 & \sigma_{2,2}^2 & 0 \\ 0 & 0 & \sigma_{3,3}^2 \end{pmatrix} \right) \tag{22}$$

and

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

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:

$$ L_i = log \frac{1}{R} \sum_{s=1}^{5} \sum_{r=1}^{R} P(p_c^s|X) \prod_{j=1}^{K} d_{ij} \left(\frac{exp(U(Y_{H^j} - TC; H^j; X, \nu))}{\sum_{k=1}^{K} exp(U(Y_{H^k} - TC); H^k; X, \nu))} \right) $$

In order to simplify the routine, I 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[26]. This is done by using the command *expand* after have constructed the dependent variable. The dependent variable is called Hc and is the choice among 15 possible alternatives[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 environment. Below the main Stata command lines 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:
forvalues r=1/50 {
gen random1_`r'=invnorm(c1_`r')
gen random2_`r'=invnorm(c2_`r')
gen random3_`r'=invnorm(c3_`r')
}
***generate the dependent variable (Hc) from the observed choices of the wife (hour_f)
***and 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
```

---

[25] A Cholesky factor is such that: $\begin{pmatrix} c_{11} & 0 \\ c_{12} & c_{22} \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{pmatrix} = \begin{pmatrix} \sigma_{1,1}^2 & \sigma_{1,2}^2 \\ \sigma_{2,1}^2 & \sigma_{2,2}^2 \end{pmatrix}$

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

[27] The choice set is defined as follow. there are 15 alternatives of hours where the couple can choose from. The choice set is defined by the cartesian product {0,10,20,30,40}x{0,40,50}

```
replace Hc=14 if hour_f==40 & hour_m==40
replace Hc=15 if hour_f==40 & hour_m==50
***now expand the dataset to get the conditional logic 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 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 double 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:
gen fc1=(Hc>=4)
gen fc2=(Hc>9)
***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 maximization 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, the variances of the random terms has to be computed[28]. Notice that for simplicity I assume that the random terms are not correlated[29]. Then the covariance matrix can be filled in and the Cholesky decomposition can be computed. After that, it 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

---

[28]What is estimated is the logarithm of the standard deviation in order to constraint the variances to be positive.

[29]I tried to estimate the covariances among these terms but they turned out to be statistically not significant.

distribution and from the multivariate normal distribution, 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. These values are then weighted for the respective probability that a particular child-care price is observed[30]. This process is carried out 50 times, which is the chosen number of draws from the wage and the random terms distribution. Given 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 per observation at the same time[31]. All this single contributions are added up into the variable L2. The last step is to pick up for each observation only the contribution that corresponds to the observed choice and to average them over all 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 etha1 etha2 etha3 etha4 etha5 etha6 sigma1 lnsig1 l11 sigma2 lnsig2 l22
sigma3 lnsig3 l33
local d "$ML_y1"
mleval 'etha1' = 'b', eq(1)
scalar mleval 'etha2' = 'b', eq(2)
scalar mleval 'etha3' = 'b', eq(3)
scalar mleval 'etha4' = 'b', eq(4)
scalar mleval 'etha5' = 'b', eq(5)
scalar mleval 'etha6' = '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
qui gen double 'L1'=0
qui gen double 'L2'=0
```

---

[30]This is done with the command *cond(x,y,z)*. In the previous section I explained that child-care expenditure can be considered only when the couple has a child under 6 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 weight is different to take into account that the probability of zero expenditure in child-care depends on the mother's working status. As 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 observation when the single contributions are added up in the subsequent line.

[31]In my Mac with 4MB of RAM and 2.2Ghz Intel core duo processor, the maximization process can last for more than 20 hours.

```
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
matrix f= (`sigma1' , 0 , 0 \ 0 , `sigma2' , 0 \ 0 , 0 , `sigma3')
capture matrix U=cholesky(f)
scalar `l11'=U[1,1]
scalar `l22'=U[2,2]
scalar `l33'=U[3,3]
forvalues r=1/50{
forvalues c=0/4{
qui gen double `random1'=random1_`r'*`l11'
qui gen double `random2'=random2_`r'*`l22'
qui gen double `random3'=random3_`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
```