



## Structural Labour Supply Models and Microsimulation

### Rolf Aaberge

Statistics Norway and Centre of Equality, Social Organization, and Performance (ESOP),  
Department of Economics, University of Oslo, Norway.  
rolf.aaberge@ssb.no

### Ugo Colombino

Department of Economics, University of Turin, Italy.  
ugo.colombino@unito.it

**ABSTRACT:** The purpose of the paper is to provide a discussion of the various approaches for accounting for labour supply responses in microsimulation models. The paper focus attention on two methodologies for modelling labour supply: (i) the discrete choice model and (ii) the random utility – random opportunities model. The paper then describes approaches to utilising these models for policy simulation in terms of producing and interpreting simulation outcomes, outlining an extensive literature of policy analyses utilising these approaches. Labour supply models are not only central for analysing behavioural labour supply responses but also for identifying optimal tax-benefit systems, given some of the challenges of the theoretical approach. Combining labour supply results with individual and social welfare functions enables the social evaluation of policy simulations. Combining welfare functions and labour supply functions, the paper discusses how to model socially optimal income taxation.

**KEYWORDS:** BEHAVIOURAL MICROSIMULATION, LABOUR SUPPLY, DISCRETE CHOICE, TAX REFORMS

**JEL classification:** C50, D10, D31, H21, H24, H31, J20

## 1 INTRODUCTION

Large microsimulation models, as originally proposed by Orcutt (1957), were meant to be behavioural. For many years, however, the microsimulation community considered behavioural responses, and in particular labour supply, either unimportant, unreliable or hard to interpret. Various motivations have progressively contributed to a more positive attitude towards the inclusion of labour supply responses into microsimulation models:

- The increasing policy interest in tax-benefit reforms, their effect on both distribution and efficiency, and the realization that policy analysis requires structural models — a long-standing message from Marschak (1953), possibly revived by Lucas (1976), in particular when the policies introduce complications and non-convexities into the opportunity sets, see (Heckman, 1974 and Hausman, 1979), and when preferences and opportunities are heterogeneous (see Aaberge, Colombino, & Strøm, 1999).
- The use of microsimulation techniques in order to compute labour supply responses, starting approximately around the early 80s (see Arrufat & Zabalza, 1986; Zabalza, Pissarides, & Barton, 1980).<sup>1</sup>
- The development of discrete choice labour supply models and of models based on various versions of the Random Utility approach (Aaberge, Dagsvik, & Strøm, 1995; Van Soest, 1995).

From around the second half of the 90s, a (cautious) introduction of labour supply responses into large microsimulation models begin. Klevmarken (1997) provides a report on early efforts towards that purpose. Creedy and Duncan (2002), Bourguignon and Spadaro (2006), Li and O'Donoghue (2013), and Aaberge and Colombino (2014) survey past and recent developments.

In Section 2 we discuss the main approach currently adopted for developing models of labour supply. Section 3 illustrates some new or alternative approaches. Section 4 addresses the issue of whether structural models are necessary and reliable. The fact that microsimulation can produce highly disaggregated and multidimensional results on the one hand contribute to the richness of the policy evaluation, on the other hand, calls for the development of synthetic indices in order to guide the comparison between alternative policies. Therefore, Section 5 is devoted to social evaluation of the simulation results and to empirical optimal taxation. Section 6 contains the conclusion and comments on future directions.

## 2 MODELLING LABOUR SUPPLY

In the same period (mid-90s) when the microsimulation community starts moving towards introducing behavioural responses, labour supply modelling benefits from an innovative research effort which had matured in the first half of the 70s, that is the random utility maximization (RUM) model developed by McFadden (1974). The crucial advantage of this approach is that the solution of the utility maximization problem is expressed in terms of comparisons of absolute values of utility rather than in terms of marginal variations of utility as in the traditional constrained utility maximization models. The RUM approach is very convenient when compared to the previous ones, since it does not require going through complicated Kuhn-Tucker conditions involving derivatives of the utility function and of the budget constraints. Therefore, it is not affected by the complexity of the rule that defines the budget set or by how many goods are contained in the utility function. Equally important, the deterministic part of the utility function can be specified in a very flexible way without worrying about the computational problems.<sup>2</sup> The most popular version adopts the “Extreme Value” distribution for the stochastic component, which leads to an easy and intuitive expression for the probability that any particular alternative is chosen (that is the multinomial or conditional logit model).

### 2.1 The discrete choice model

This approach essentially consists in representing the budget set with a set of discrete alternatives or jobs. Early and path-breaking contributions include Zabalza et al. (1980), where labour supply is represented in terms of probabilities of choosing among alternative hours of work or alternative jobs. This contribution, however, is essentially an ordinal probit analysis. Especially in view of modelling simultaneous decisions on the part of household partners, the Conditional Multinomial Logit model appears much more convenient. This is the line chosen by Van Soest (1995). Although this very influential contribution can be classified as belonging to the RUM family, we denote it more specifically as a Discrete Choice (DC) model, because: (i) the discreteness of the opportunity set is a distinctive feature of it (this is not the case in general for RUM models); (ii) the random term that generates the probabilistic choices is given an eclectic interpretation that includes both the RUM-McFadden (McFadden, 1974; 1984) interpretation and the optimization error interpretation (the latter leading to a non-random utility model). Besides Van Soest (1995), many contributions have adopted the DC model during the last two decades.

The DC model typically treats also couples with simultaneous decisions of the two partners, but in order to keep the illustration simple, we will discuss the single case below: the extension to couples is straightforward. The household chooses among  $T+1$  alternatives or  $b = 0, 1, \dots, T$ . The utility is first defined as non-stochastic,  $v(f(wh, I), h)$ , where  $w$  is the fixed (individual-specific) gross wage rate,  $I$  is the exogenous income and  $f(\dots)$  is the tax-transfer rule that transforms gross incomes into net available income. In order to model the observed hours of work as the result of a probabilistic process, a random variable  $\varepsilon$  is added to the previously defined utility function:  $v(f(wh, I), h) + \varepsilon$ . As mentioned above, the random term is typically given two different interpretations (for instance Van Soest, 1995): (i) the utility contribution of unobserved characteristics of the alternative choices; (ii) a measurement/optimization error. Interpretation (i) is compatible with the classic RUM interpretation and implies that the household are observed as choosing exactly what they prefer, and what they prefer is decided on the basis of  $v(f(wh, I), h) + \varepsilon$ . Interpretation (ii) instead implies that the household's preferences are measured by  $v(f(wh, I), h)$  but the alternative to which they are matched does not maximize  $v(f(wh, I), h)$  but rather  $v(f(wh, I), h) + \varepsilon$ : this might happen because they make errors or because some other unexpected process displaces them from the preferred choices. However, the two interpretations in principle have very different implications in view of the simulation and of the welfare evaluation. The contributions adopting the DC approach stress the importance of a very flexible specification of  $v(f(wh, I), h)$  and of checking for its quasi-concavity (for instance Van Soest, 1995). This focus of attention suggests that this approach indeed tends to consider  $v(f(wh, I), h)$  as the true utility function and  $\varepsilon$  as a measurement/optimization error.<sup>3</sup> Consistently, preference heterogeneity is preferably introduced through random preference parameters.

The household is assumed to choose  $b$  so as to maximize  $v(f(wh, I), h) + \varepsilon$ . By assuming that  $\varepsilon$  is i.i.d. Type I Extreme Value, one gets the Multinomial Logit or Conditional Logit expression for the probability that the household is observed working  $b$  hours:<sup>4</sup>

$$P(h) = \frac{\exp\{v(f(wh, I), h)\}}{\sum_{y=0}^T \exp\{v(f(wy, I), y)\}} \quad (1)$$

Equation 1 usually does not fit labour supply data very well. Van Soest (1995) notes that the model over-predicts the number of people working part-time. More generally, certain types of jobs might differ according to a number of systematic factors that are not accounted for by the observed variables contained in  $v$ : (a) availability or density of job-types; (b) fixed costs; (c) search costs; (d)

systematic utility components. In order to account for these factors, the following “dummies refinement” can be adopted. Let us define subsets  $S_0, \dots, S_L$  of the set  $(0, 1, \dots, T)$ . Clearly, the definition of the subsets should reflect some hypothesis upon the differences between the values of  $b$  with respect to the factors (a) – (b) mentioned above. Now we specify the choice probability as follows (Equation 2):

$$P(h) = \frac{\exp\left\{v(f(wh, I), h) + \sum_{\ell} \gamma_{\ell} 1(h \in S_{\ell})\right\}}{\sum_{y=0}^T \exp\left\{v(f(wy, I), y) + \sum_{\ell} \gamma_{\ell} 1(y \in S_{\ell})\right\}} \quad (2)$$

where  $1(e) = 1$  if  $e$  is true. Many papers have adopted this refinement, for instance Van Soest (1995) and Kalb (2000) among others. Aaberge et al. (1995, 1999) and Colombino (2013) also implement a similar procedure, which however is based on a specific structural interpretation of the dummies and of their coefficients. An alternative adjustment consists of imputing a monetary cost (or benefit) to some ranges of work hours:

$$P(h) = \frac{\exp\left\{v(f(wh, I) + \sum_{\ell} c_{\ell} 1(h \in S_{\ell}), h)\right\}}{\sum_{y=0}^T \exp\left\{v(f(wy, I) + \sum_{\ell} c_{\ell} 1(y \in S_{\ell}), y)\right\}} \quad (3)$$

A popular specification of Equation 3 is interpreted as accounting for fixed costs  $c$  of working — for instance see the survey by Blundell, Chiappori, Magnac, and Meghir (2007).

## 2.2 The random utility – random opportunities model

The Random Utility – Random Opportunities (RURO) model is an extension of McFadden’s RUM model. The utility is assumed to be of the following form (Equation 4):

$$U(f(wh, I), h, j) = v(f(wh, I), h) + \varepsilon(w, h, j) \quad (4)$$

where  $h$  is hours of work,  $w$  is the wage rate,  $I$  is the exogenous income,  $f$  is a tax-transfer function that transforms gross incomes into net income,  $j$  is a variable that captures other job and/or individual characteristics and  $\varepsilon$  is a random variable that varies across market and non-market alternatives.

A first difference with respect to the DC model is that the utility function is directly specified as stochastic. The random component is interpreted as in the presentations of the Conditional Logit model by McFadden (1974): besides the observed characteristics, there are other characteristics  $j$  of the job or of the household-job match that are observed by the household but not by the econometrician. Commuting time or required skill (when not observed by the analyst) are possible examples of the characteristics captured by  $j$ . Their effect upon utility is captured by  $\varepsilon(w, h, j)$ .

Second, the households maximize their utility by choosing not simply hours but rather opportunities (“jobs”) defined by hours of work  $h$ , wage rates  $w$  (which can change across jobs for the same household) and other unobserved (by the analyst) attributes  $j$ . In the DC model, the households’ choices (how many hours of work) are analogous to the choices of a consumer deciding how many units of a consumption good (like meat, milk or gasoline) to buy every week. In the RURO model, the household is closer to the McFadden’s commuter choosing among car, train or the BART shuttle when travelling along the San Francisco Bay (Domencich & McFadden, 1975).

Third, besides not observing the other job characteristics  $j$ , the analyst does not know exactly which and how many jobs are contained in the household opportunity set; therefore, the opportunity set can be seen as random from the analyst’s viewpoint. The opportunity set will in general contain more than one job of the same  $(w, h)$  type. These jobs will differ depending on the value of other, by the analyst, unobserved attributes. This implies that the number, or the density, of jobs belonging to the different types will play a crucial role in the model.

In Aaberge et al. (1995) the range of values of  $(w, h)$  is assumed to be continuous. Let  $B$  be the set of admissible values of  $(w, h)$  and  $p(x, y)$  the density of jobs of type  $(x, y)$ . The household chooses  $h$  and  $j$  so as to maximize  $v(f(wh, I), h) + \varepsilon(j)$ . Then it turns out that we get the (continuous) conditional logit expression for the probability density function of a  $(w, h)$  choice:

$$\varphi(w, h) = \frac{\exp\{v(f(wh, I), h)\} p(w, h)}{\int_{(x, y) \in B} \exp\{v(f(xy, I), y)\} p(x, y) dx dy} \quad (5)$$

Equation 5 is based on Dagsvik (1994). The model is close to the continuous spatial model developed by Ben-Akiva and Watanatada (1981). It can also be seen as an extension of the McFadden’s Conditional Logit model where the systematic utility of a job type  $(w, h)$  is “weighted”

by the number of jobs of that type available in the opportunity set. Aaberge et al. (1999) provide a transparent and simple proof for a discrete version of Equation 5:

$$\varphi(w, h) = \frac{\exp\{v(f(wh, I), h)\} p(w, h)}{\sum_{(x, y) \in B} \exp\{v(f(xy, I), y)\} p(x, y)} \quad (6)$$

The discrete version can be interpreted either as a more realistic representation or as computational simplification of the continuous version.

So far, in all the applications of the RURO the opportunity density  $p(w, h)$  is first factorized as (Equation 7):

$$p(w, h) = \begin{cases} p_1 g_1(h) g_2(w) & \text{if } h > 0 \\ 1 - p_1 & \text{if } h = 0 \end{cases} \quad (7)$$

where  $p_1$  denotes the density of alternatives with  $h > 0$ , that is market jobs,  $g_1(h)$  and  $g_2(w)$  are the densities of  $w$  and  $h$  conditional on  $h > 0$ . The conditional density of hours is specified as uniform with peaks (to be estimated) corresponding to part-time and full-time. The conditional density of the wage rates is assumed to be log-normal. Details can be found in Aaberge et al. (1999, 1995). All the densities  $p_1, g_1(h), g_2(w)$  and the density of  $w$  can depend on household or job characteristics.

By looking at Equation 6, we can see that the solution of the utility maximization problem is expressed in terms of comparisons of absolute values of utility rather than in terms of marginal variations of utility and it is not affected by the specification of  $v(.,.)$  or  $f(.,.)$ . One can choose relatively general and complicated specifications for  $v$  and/or accounting for complex tax-transfer rules  $f$  without affecting the characterization of behaviour and without significantly affecting the computational burden involved by the estimation or simulation of the model. This holds for both the discrete and the continuous version of the model. It is not often realized in the literature that the advantages of RUM or of RURO are due to the representation of choice as the maximization of a random utility, rather than to the discreteness of the choice set.

Note that Equation 1 can be seen as a special case of Equation 6 when the wage rate  $w$  is treated as a fixed characteristic of the household (invariant with respect to the alternatives) and  $p(x, y) =$  constant for all  $(x, y)$ .

It is also useful to observe that the opportunity density  $p(x,y)$  can be specified in such a way that Equation 6 reduces to a DC model with dummies refinement. For example, Colombino (2013) starts by considering a model with fixed individual specific wage rates (Equation 8):

$$\varphi(h) = \frac{\exp\{v(f(wh, I), y)\} p(h)}{\sum_{y \in B} \exp\{v(f(wy, I), y)\} p(y)} \quad (8)$$

By specifying the opportunity density  $p(y)$  as uniform-with-peaks, we get the following expression (Equation 9):

$$\varphi(h) = \frac{\exp\{v(f(wh, I), h) + \gamma_0 \mathbf{1}(h > 0) + \sum_{\ell=1}^L \gamma_\ell \mathbf{1}(h \in S_\ell)\}}{\sum_{y \in B} \exp\left\{v(f(wy, I), x) + \gamma_0 \mathbf{1}(y > 0) + \sum_{\ell=1}^L \gamma_\ell \mathbf{1}(y \in S_\ell)\right\}} \quad (9)$$

with

$$\gamma_0 = \ln J + A_0, \gamma_\ell = \ln\left(\frac{J_\ell}{J}\right) + A_\ell \quad (10)$$

where  $J$  = number of alternatives with  $h > 0$ , and  $J_\ell$  = number of alternatives with  $h \in S_\ell$  (for instance,  $S_\ell$  might be the set of hours values classified as “part-time”) and  $A_0$  and  $A_\ell$  are constants (Equation 10). Equation 9 is formally equivalent to the DC model with the “dummies refinement”. However, here the coefficients  $\gamma$  have a specific structural interpretation, which —as we will see in the section dedicated to policy simulation— can be used to develop an equilibrium simulation procedure.

### 2.3 The representation of the opportunity set

In the continuous version of the RURO model, the opportunity set in principle can contain the whole positive quadrant, that is all the positive values of  $(w, b)$ . If instead a discrete representation of the choice set (as in the DC model or as in the version, see Equation 6, of the RURO model) is adopted, then one has to decide which alternatives are to be included in the opportunity set (besides the chosen alternative). DC models typically assume the opportunity set is fixed and imputed to every household. For example, one might divide the hour interval  $(0, T)$  into equal sub-intervals and pick one value in each sub-interval (for instance the midpoint, or a randomly chosen point). The wage rate is also fixed and household-specific: therefore, for every value  $b$ , the corresponding



gross earnings are equal to  $wb$ . In the RURO models, the opportunity set is unknown since the opportunity density  $p(w,b)$  must be estimated. The opportunity set used in the estimation (and in the simulations) can then be interpreted as a sample drawn from an unknown population. Therefore, the sampling method emerges as a relevant issue. Aaberge et al. (1999, 1995) sample alternative  $(w,b)$  values from a pre-estimated density  $q(w,b)$  and, following Ben-Akiva and Lerman (1985), re-weight Equation 6 as follows:

$$\varphi(w, h) = \frac{\exp\{v(f(wh, I), h)\} p(w, h)/q(w, h)}{\sum_{(x, y) \in \hat{B}} \exp\{v(f(xy, I), y)\} p(x, y)/q(x, y)} \quad (11)$$

where  $\hat{B}$  is the sample of market and non-market alternatives (Equation 11). Aaberge, Colombino, and Wennemo (2009) discuss and evaluate different methods of representing the opportunity set and find that they might have an important impact on the results of the policy simulation.

## 2.4 Unobserved wage rates

The problem of unobserved wage rates for those who are not working can be solved either with a simultaneous procedure or with a two-step procedure. When adopting a simultaneous estimation with a DC model, one should also treat the wage rate  $w$  as an endogenous outcome and account for the fact that  $w$  is not observed for the non-workers in the sample. For that purpose, we must specify a probability density function  $m(w)$ . Starting from Equation 1, the likelihood of an observation with non-zero hours  $b$  and wage rate  $w$  would then be (Equation 12):

$$P(w, h) = \frac{\exp\{v(f(wh, I), h)\} m(w)}{\sum_{k=0}^T \exp\{v(f(wk, I), k)\}} \quad (12)$$

The likelihood of an observation with  $b = 0$  and unobserved wage rate would instead be (Equation 13):

$$P(h = 0) = \int \frac{\exp\{v(f(0, I), 0)\} m(w) dw}{\sum_{k=0}^T \exp\{v(f(wk, I), k)\}} \quad (13)$$

In RURO models, the wage rate is endogenous from the very start. Therefore (in the continuous version), the likelihood of a choice  $(w, b)$  is given by Equation 6. By inserting Equation 7 into Equation 6 we get:

$$\varphi(w, h) = \begin{cases} \frac{\exp\{v(f(wh, I), h)\} p_1 g_1(h) g_2(w)}{\exp\{v(f(0, I), 0)\} (1 - p_1) + \sum_{(x,y) \neq 0} \exp\{v(f(xy, I), y)\} p_1 g_1(y) g_2(x) dx dy}, & h > 0 \\ \frac{\exp\{v(f(0, I), 0)\} (1 - p_1)}{\exp\{v(f(0, I), 0)\} (1 - p_1) + \sum_{(x,y) \neq 0} \exp\{v(f(xy, I), y)\} p_1 g_1(y) g_2(x) dx dy}, & h = 0 \end{cases} \quad (14)$$

Alternatively, one could use a two-step procedure for imputing unobserved wages. In the first step, the wage equation is estimated. In the second step, the predicted wage rate replaces the missing values (or, alternatively, both the missing and the observed values). The random term of the wage equation is added to the systematic part and integrated (or “averaged”) out with a simulation procedure (see for instance Van Soest, 1995). Both the simultaneous and the two-steps procedures illustrated above assume that the random term of the wage equation is uncorrelated with the random term of the utility function. However, one might want to allow for a correlation of the wage rate random component with one or more random parameters of  $v(f(wh, I), h)$  due, for example, to a dependence of the wage rate on previous decisions (see for instance Blundell & Shephard, 2012; Gong & van Soest, 2002).

## 2.5 Unemployment

In RURO models,  $\varepsilon$  is interpreted as part of the utility function and therefore  $h = 0$  is an optimal choice. Involuntary unemployment can be considered in different ways depending on which interpretation of which concept of involuntary unemployment is adopted. A first interpretation is associated with the opportunity set. An individual is assumed to be involuntary unemployed if the set of available market opportunities is empty, or contains “too few” elements, or elements with “two poor” characteristics (for instance low wage rates, bad non-pecuniary features, etc.). The qualification of “involuntary” is motivated by the exogeneity of an “unattractive” opportunity set. The opportunity density  $p(w, h)$  in general allows for this possibility. A second interpretation sees involuntary unemployment as an unanticipated displacement from the chosen alternative. The most natural way to implement this interpretation would be to complement the basic labour supply model with an exogenous latent index equation (see for instance Blundell et al., 2007). As a matter of fact, this approach has been adopted so far with DC models but not with RURO models.

If  $\mathcal{E}$  is interpreted as an optimization error rather than as part of the utility— as is more common with DC models— then some of the observations with  $h = 0$  might be interpreted as involuntary

unemployed. The idea here is that the individual maximizes (by mistake)  $U + \varepsilon$  rather than the true utility  $U$ . Maybe the involuntary unemployed could be identified as those with  $h = 0$  and systematic utility sufficiently close (in some sense) to the systematic utility of those with  $h > 0$ . To the best of our knowledge, this line of research has not been pursued. Alternatively, one could interpret the optimization errors due to  $\varepsilon$  as accounting for more modest displacements such as underemployment or overemployment and instead model unemployment with a latent index equation (Blundell et al., 2007).

## 2.6 Generalizations

Both the DC and the RURO model can be easily generalized to include several dimensions of choice. Besides simultaneous decisions on the part of partners in a couple, one might include other decisions such as: labour supply of other members of the household, consumption of goods and services, fertility, choice of child-care mode, sector of employment, other dimensions of labour supply (occupational choice, educational choices, job search activities, etc.). For example, Aaberge and Colombino (2013) and Dagsvik, Locatelli, and Strøm (2009) include the choice between employment in the private sector and the public sector; Kornstad and Thoresen (2007) model the simultaneous choice of labour supply and child-care; Haan and Wrohlich, (2011) analyse fertility and employment, Hoynes (1996) and Aaberge and Flood (2013) analyse labour supply and welfare participation.

A potential limitation of the RUM models based on the independent and identical extreme value distribution for the random component  $\varepsilon$  is the IIA assumption, which in turn implies restrictions on the behavioural responses (see Ben-Akiva & Lerman, 1985). Some contributions have opted for alternative distributional assumptions (see Keane & Moffitt, 1998). However, advances with simulation-based methods (Train, 2003), have made it feasible to overcome this limitation by assuming GEV (generalized extreme value) distributions (for instance Nested Logit models) or random parameters, while preserving the main convenient analytical advantages of the extreme value distributions. By assuming that one or more preference parameters are stochastic one gets the so-called Mixed Logit model (McFadden & Train, 2000).

### 3 NEW DEVELOPMENTS AND ALTERNATIVE MODELS

We mention here three important research strands that have been developed during the last decades, either as refinements of the standard labour supply model or as innovative or alternative approaches.

1. Stochastic dynamic programming (SDP) model (see for instance Keane, 2011; Wolpin, 1996)

There are various motivations for using SDP models. First, many choices —notably human capital decisions, occupational choices, fertility, etc.— have important intertemporal implications, namely the effects of decisions taken today have important effects in the future. Second, many policies have an intrinsic intertemporal dimension, for instance there might be time limits, or it might be that the amount of services I decide to get today affects the amount of services I can get tomorrow (Swann, 2005). Third, an important source of uncertainty in current decisions is the expectation of future changes in policies, e.g. expectations on whether a certain policy is temporary or permanent (Keane & Wolpin, 2002a, 2002b).

2. Non-unitary models of household behaviour

Where the household is not represented as a fictitious individual but rather as a set of individuals who, somehow, arrive at a collective decision. A major aim is developing models that can analyse intra-household allocation of resources (for instance among genders) and the effects of policies upon different members of the households. As to the way of modelling the process that leads to the collective decision, there are two main lines of research: (i) the “sharing rule” approach (see for instance Bloemen, 2010; Chiappori, 1988). Here, the intra-household allocation process is given a reduced form representation. This way of proceeding requires minimal *a-priori* assumptions (namely, the household attains, somehow, a Pareto-efficient allocation), but in principle makes the model not applicable to *ex-ante* policy evaluation, unless one is prepared to assume that the “sharing rule” is policy-invariant; (ii) the explicit structural representation of intra-household allocation process. For example, McElroy and Horney (1981) have proposed Nash bargaining. So far, this second approach has been much less popular than the “sharing rule” one, although its structural character makes it more promising in view of policy simulation (see for instance Del Boca & Flinn, 2012).

### 3. The “taxable income” approach

This is especially relevant for applications in public finance and optimal taxation. As a matter of fact, labour supply has many dimensions: not only hours of work, but also search, occupational choice, training, “effort”, etc. Although there might be a specific interest in modelling all these choices, from the public finance perspective what is mostly relevant is their combined effect, i.e. the amount of taxable income. Feldstein (1995) argues that for the purpose of measuring the efficiency effect of (marginal) tax reforms, it is sufficient to have an estimate of the elasticity of taxable income with respect to the tax rates. The argument sounds attractive since an estimate of the taxable income elasticity is relatively easy to obtain and furthermore the data on taxable income might be more reliable than data upon the various dimensions of labour supply (hours, etc.). If we denote taxable income with  $z$  an implication is that the reference model becomes  $\max u(c, z)$  s.t.  $c = f(z)$  rather than the standard framework  $\max u(c, h)$  s.t.  $c = f(wh, I)$ , where  $f()$  denotes the tax-transfer rule that transforms taxable income(s) into net available income. The taxable income approach tends to be taken as a partner of the non-structural approach (and therefore appropriate only for the evaluation of marginal reforms), but in principle nothing prevents to adopt with a structural model. Chetty (2009) provides a discussion of the conditions under which the argument of Feldstein (1995) is valid and of its implications for empirical research.

## 4 EVALUATION OF STRUCTURAL MODELS

Many authors have raised doubts upon the reliability of structural models as compared with the (supposed) robustness of evidence produced by (*ex-post*) experimental or quasi-experimental analysis. In view of *ex-ante* policy evaluation, the issue is twofold: (i) are there alternatives to structural models?, and (ii) how do we evaluate structural models and how do they compare with other approaches?

When answering question (i) one must carefully distinguish between type of data and type of models (or parameters) to be estimated. Often, we observe a tendency to associate structural models with observational data and ex-post program evaluation with experimental or quasi-experimental data. Although this is what goes on in most cases, in principle nothing prevents the use of experimental or quasi-experimental data for the estimation of structural models (for instance Bargain & Doorley, 2017; McElroy & Horney, 1981). A second possible source of confusion comes from erroneously associating structural modelling with the use of “convenient” parametric functional forms: although this might be a common practice, most of the research done on “non-

parametric” (or “flexible”) estimation addressed to policy evaluation is structural (for instance Blomquist & Newey, 2002; Matzkin, 2013). A third erroneous perception consists of identifying structural models with models based on utility maximization. Again, while utility maximization is the “mainstream”, most of the “agent-based” approach is structural.<sup>5</sup> What counts in view of *ex-ante* evaluation is that a set of relevant parameters (or primitives) be identified as policy independent. Depending on the class of policies we are interested in different sets or combinations of parameters might be sufficient for the purpose (Marschak, 1953). Of course, the point is that data, *by themselves*, whether experimental or quasi-experimental or non-experimental, are not sufficient to identify policy-invariant parameters. Therefore, the answer to question (i) is negative: *ex-ante* evaluation requires a structural model, whether parametric or non-parametric, based on utility maximization or not, explicit or implicit, estimated on observational or (quasi-) experimental data, etc.<sup>6</sup> Let us turn to question (ii). The structural econometric community tends now to see models as approximations. Ordinary statistical testing is informative on the precision of the parameter estimates of the model but less so on how useful the estimated model is. This pragmatic approach would seem to entail a shift of focus from the issue of identification to the issues of external validation and out-of-sample prediction (Keane, 2010). The amount of out-of-sample testing so far is limited (for instance Aaberge et al., 2009; Aaberge & Flood, 2013; Keane & Moffitt, 1998; Keane & Wolpin, 2002a, 2002b) but reassuring. A supplementary evidence provided by out-of-sample prediction exercises suggests that flexible a-theoretical models —as compared with structural models— tend to perform better in-sample but worse out-of-sample.

## 5 POLICY SIMULATION

We start by asking, when is information on behavioural responses needed? Non-behavioural simulations may be sufficiently informative provided the policy changes or the reforms can be represented as marginal changes in net wages and/or in unearned income. Let  $V(w, I)$  be the indirect utility function, where  $w$  is the net wage rate and  $I$  is the unearned income. Let us suppose that the reform can be represented as a marginal change  $(dw, dI)$ . Then we have:

$$dV = \frac{\partial V}{\partial w} dw + \mu dI,$$

where

$$\mu \equiv \frac{\partial V}{\partial I}$$

is the marginal utility of income. By applying Roy's theorem, we get:

$$\frac{dV}{\mu} = hdw + dI.$$

The right-hand side is the change in the budget, conditional on the pre-reform labour supply  $b$ . The left-hand side is the monetary equivalent of the change in utility. Therefore, the result tells us that the change in the budget (that is the basic result produced by a non-behavioural simulation) is a money-metric measure of the change in utility. Similar arguments can be generalized so that a non-behavioural simulation can be complemented by point-estimates of elasticities or other local measures of behavioural responses (Chetty, 2009).

When the reforms involve non-marginal changes in the budget constraint, we typically want a prediction of the new choices, in particular of the new value of  $b$  or some function of it. With DP or RURO models, we can choose between two alternative procedures:<sup>7</sup>

- (i) Compute the expected chosen value of the variable of interest, based upon the estimated choice probabilities (see for instance Colombino, 2013).
- (ii) Simulate the value of the systematic utility and of the random component corresponding to each alternative in the opportunity set. Identify the alternative with the highest utility and compute the corresponding value of the variable of interest. Typically, the random components are kept fixed across the different policy regimes that one might want to simulate and compare. When comparing a reform to the current system, it is appropriate to simulate the latter as well. The simulated current system, although not identical but reasonably close to the observed one, will provide a consistent basis for the comparison.

### 5.1 Short-run, long-run, comparative statics

The results of non-behavioural policy microsimulation are usually interpreted as predictions of the very short term, when agents and market interactions did not have time yet to adjust to the new policy. Even in the long-run, non-behavioural results might be informative enough, provided the reforms can be represented as marginal changes in the budget constraint. The interpretation of behavioural microsimulation results raises more controversial issues. The typical policy simulation exercise computes the labour supply effects while leaving the wage rates unchanged. Some authors (see for instance Creedy & Duncan, 2005) interpret this scenario as the "month after" prediction, with households making new choices but the market mechanisms is still late in the process of adjusting wage rates, labour demand, etc. In our view, however, the appropriate approach with

static behavioural microsimulation models is comparative statics, that is, we want to compare two different equilibria induced by two different policies. With the notion of equilibrium, we refer in general to a scenario in which the economic agents make optimal choices (that is, they choose the best alternative among those available in the opportunity set) and their choices are mutually consistent or feasible. Creedy and Duncan (2005) and Peichl & Siegloch (2012) have proposed procedures where DC labour supply models (as defined in Section 2) are complemented by a function of labour demand and the wage rates are adjusted so that the market attains the equilibrium. With RURO models a different procedure must be used, since their specification already includes a representation of the labour demand side (that is the density of available market jobs). Since a reform in general will induce a change in labour supply, it follows that in equilibrium also the number of available jobs will have to change. Colombino (2013) proposes and exemplifies an iterative simulation procedure that exploits the structural interpretation of the coefficients of the alternative-specific constants given in Equation of Section 2.2.

## 5.2 Evolution of labour supply elasticities

Although wage and income elasticities cannot be considered as autonomous parameters they provide useful information of the potential for stimulating labour supply by appropriate policy reforms. The comparability of the elasticities found in the literature has, however, been questioned due to differences in data and choice of modelling framework. To account for the effect of data and methodological differences, Bargain, Orsini, and Peichl (2014) assessed labour supply elasticities for 17 EU countries and the US on the basis of harmonized data covering a restricted period (1998–2005) and by using the same version of the random utility model (RUM) as previously has been used by Van Soest (1995) for instance. Although the RUM, as discussed above, suffers from certain shortcomings as compared to the RURO the use of a unified framework will nevertheless improve the cross-country comparability of the derived labour supply elasticities. The results provided by Bargain et al. (2014) suggest that the large variation in previously published elasticities is mainly due to differences in modelling framework and different observation periods of the data. Thus, one might question whether the sharp decline in labour supply elasticities in Europe and the US is due to differences in measurement method and methodological framework? A crucial change in methodological approach took, as indicated above, place in the mid-90s when the approach introduced by Hausman (1979) was replaced by various versions of the random utility model. By using elasticities derived from these two modelling frameworks, Bargain and Peichl (2016) suggest that elasticities actually have declined since the 1980s. Below we will discuss this



claim on the basis of elasticities derived from estimates of the RURO model for Norway in 1979, 1986, 1994, 2006 and 2011.

During the period 1979–2011, the effect of a wage increase on total labour supply in Norway changed from being positive to become almost zero (Bhuller, Vestad, & Aaberge, 2016). While in earlier years a wage increase led to a significant increase in overall labour supply, it will now lead to almost no change in labour supply. This trend is as expected and mostly due to an increase in education and a formidable real wage growth over the past 35 years. A significantly larger proportion of married women was employed in 2011 than in 1979, which means that the potential for further increase in employment has significantly decreased. Greater degree of equality in education among women and men and generous parental leave plans have also contributed to the fact that the fathers have taken parental leave from work and become more involved in the service production at home, which might have contributed to more equal labour supply behaviour for females and males over time. Increased weight on leisure today than 35-40 years ago is due to the income effects from economic growth and a doubling in household incomes over the last 35 years. For those who already live in Norway, it may therefore prove to be challenging to maintain the current level of employment in a future with continued economic growth if the trend of a greater appreciation of leisure continues.

Bhuller et al. (2016) found for 2011 that individuals with low income and few hours worked responded more strongly to a wage increase than those with high income and many hours worked. This is largely because low income individuals have a greater potential to increase their labour supply, but it also relates to the fact that low income individuals generally have the least attractive jobs in terms of hourly wage and job content. Therefore, economic stimulation will have a stronger effect on offered jobs for people with low incomes than for high income people. This relationship has been found based on Norwegian data for all years that is 1979, 1986, 1994, 2006 and 2011. A similar relationship is also found for two sets of data for Sweden (Aaberge & Flood, 2013; Aaberge, Colombino, & Strøm, 2000) and two datasets for Italy (Aaberge et al., 2000; Aaberge, Colombino, & Strøm, 2004).

As indicated above, labour supply behaviour for women and men has become more similar over time, although the elasticities of married women from immigrant groups are significantly higher than for men and immigration increased significantly over the recent 15 years. This must be seen in the context that many women from immigrant groups are not in work or work for a few hours and therefore have great potential for increasing employment, while most women from the rest of

the population work full-time or long part-time. Thus, since observed participation in the labour market is significantly lower among immigrant groups than among ethnic Norwegians they have a larger potential for increasing labour supply. By decomposing the overall elasticities by participation and hours elasticities, hours responses for married/cohabitants were found to be more affected by changes in hourly wages than the decision to work, some of which belong to the non-immigrant population. For immigrants, the picture is more complex, and the results vary with immigrant background. However, regardless of immigrant background, the pattern of the elasticities is like those for ethnic Norwegians and show to be relatively high for the lowest income deciles, and then they fall significantly with the income size. This is also consistent with results in studies based on data from the 70s, 80s and 90s. While the high-paid previously had small positive wage elasticities, they have negative wage elasticities in 2006 and 2011. Income elasticities for all immigrant groups except those from Western Europe, North America, and Oceania are higher than for the rest of the population and are run by both participation and working time decisions.

### 5.3 Labour supply simulations addressing specific policies

In this section we focus on three specific applications that in the last decades attracted much attention: (i) in-work benefits or tax credits, (ii) basic income, and (iii) the flat tax. The first two policies are part of the debate on redesigning the welfare system, the last one is a recurrent idea aiming at simplification and efficiency.<sup>8</sup>

Since the end of the Second World War, means-tested transfers with phasing out rate close to 100 per cent —a form of Conditional Basic Income (CBI)— prevailed as the main form of income support mechanism in most Western countries. This policy introduces a disincentive to work, especially so for people with a low wage rate, together with further problems: high administration costs, “welfare stigma” effects and take-up costs leading to low take-up rates, incentives to under-reporting of income, errors in applying eligibility criteria and litigation costs (see for instance Atkinson, 2015; Friedman, 1962). Also as a response to these problems, the so-called Negative Income Tax (NIT) was proposed by Friedman (1962) and supported by many economists.<sup>9</sup> Since the second half of the 70s, in many countries, various reforms of the income-support policies have taken still a different path: work-fare programs, less generous transfers, policies targeted towards smaller segments of the population, a more sophisticated design of eligibility conditions and of the timing of transfers, in-work benefits or tax credits in order to strengthen the incentives to work (Moffitt, 2003). The design of the various tax credit systems varies along many dimensions, where the Swedish and the US versions have represented the extremities. As opposed to the US system

the 2007 Swedish system was universal and not phased out and thus reduced taxes for all working individuals at all earnings levels. By contrast, since the EITC (earned income tax credit) system in the US is phased out at a moderate earnings level and targeted to low-income families, redistributive concerns appear to be a major justification for its design. Evaluations of phase-out and non-phase-out versions of the tax credit system have been carried out by Bhuller et al. (2016) and Aaberge and Flood (2013) based on Norwegian and Swedish data. As expected, the phase-out versions generate lower labour supply responses, lower budget deficit and larger decrease in income inequality compared to the non-phased-out systems. More recently, in many countries a new interest is emerging for a still different reform direction: less conditioning and simpler designs closer to the original Friedman (1962) proposal of the NIT, with Unconditional Basic Income (UBI) as a limit case with no means-testing (for instance Atkinson, 2015; Van Parijs, 1995).<sup>10</sup> Also the so-called Flat-Tax (FT) —as the NIT or the UBI— is an idea pointing towards simplification and is often associated with NIT-like mechanisms (see for instance Atkinson, 1996). The likely effects on labour supply of these policies are an important issue for their evaluation. The FT has been analysed with behavioural microsimulation models (for instance by Aaberge et al., 2000; Fuest, Peichl, & Schaefer, 2008; Paulus & Peichl, 2008; Peichl, 2014) provides a recent survey. UBI and other members of the NIT class have also been analysed with different results (see for instance Aaberge et al., 2000, 2004; Clavet, Duclos, & Lacroix, 2013; Colombino, 2015; Colombino & Narazani, 2013; Horstschräer, Clauss, & Schnabel, 2010; Islam & Colombino, 2018; Jessen, Rostam-Afschar, & Steiner, 2017; Scutella, 2004; Sommer, 2016). Islam and Colombino (2018) examine in various European countries the case for an optimal tax-transfer rule in the class NIT+FT, assuming all incomes are treated according to the same rule. They find that the current system is always dominated (social welfare wise) by at least one member of the class NIT+FT. Labour supply effects are small but not irrelevant. In most cases UBI is preferred to CBI, the latter inducing more “welfare dependence”. It might be the case that the important effects would come from changes in administration costs; most likely a reduction when adopting policies with simpler designs. So far, however, structural models and microsimulation procedures have not been able to account for the implications of administration costs. A gap which is to be filled in future work.

#### 5.4 Optimal taxes

Optimal taxation theory addresses the question of how tax-transfers rules should be designed to maximise a social welfare function subject to the public revenue constraint and considering that households choose labour supply, or more generally “effort”, in order to maximize their utility function subject to the budget constraint defined by the tax-transfer rule. Mirrlees (1971) is the

path breaking theoretical contribution. The studies linking theoretical optimal taxation to empirical research and policy analysis proceed as follows. The researcher looks for an analytical solution to the optimal taxation problem, that is a “formula” that allows to compute the optimal tax design as function of observed variables and parameters. For example, using a simplified version of Saez (2001) —assuming identical preferences, no income effects and interior solutions— the following result is obtained:

$$\frac{T'(z)}{1-T'(z)} = \left( \frac{1}{e(z)} \right) \left( \frac{1-F(z)}{zf(z)} \right) (1-G(z)) \quad (15)$$

where  $T'(z)$  is the marginal tax rate applied at (taxable) income  $z$ ,  $e(z)$  is the elasticity of  $z$  with respect to  $1 - T'(z)$ <sup>11</sup>,  $F(z)$  and  $f(z)$  are the distribution function and the density function of  $z$  and  $G(z)$  is a relative social weight attached to individuals with income greater than  $z$ . Note that this formulation adopts the “taxable income” approach (see Section 3), rather than the more traditional labour supply approach. Of course, Equation 15 is not a direct solution, since  $z$  depends on the tax rule  $T(\cdot)$ . Therefore, in order to compute  $T'(\cdot)$ , we must specify a structural model that explains how  $z$  depends on  $T(\cdot)$ , (see for instance Brewer, Saez, & Shephard, 2008), and impute (based on external estimates, guesses, calibrations or just assumptions)  $e(\cdot)$ ,  $F(\cdot)$ ,  $f(\cdot)$  and  $G(\cdot)$ . Mirrlees (1971) and Saez (2001), among others, using equations similar to Equation 15, or more general formulations with income effects, to get an optimal tax profiles that is pretty close to a FT with a lump-sum transfer for low incomes. Tuomala (2010), however, shows that the results are very sensible to the assumptions upon preferences and productivity distribution. Saez (2002) adopts a discrete choice framework that accounts for both intensive and extensive responses, with results that suggest the possible optimality of in-work benefits (rather than lump-sum transfers) policies for low income households. More recent contributions argue also in favour of progressive taxation and high top marginal tax rates (see for instance Diamond & Saez, 2011).

The role of elasticity, or elasticities, of labour supply is central in this literature. This is evident in Equation 15 and carries over to more general formulations were both intensive— and extensive— margin elasticities are present and can depend in general on the level of income. The early contributions mainly imputed alternative values using elasticity as a tool for sensitivity analysis. More recent contributions use microeconomic estimates. The influential work of Saez (2001) elaborates upon the possibility of computing optimal taxes only based on estimated elasticities without a structural labour supply model. The idea has been generalized by Chetty (2009) and labelled as the “sufficient statistics” approach, which in fact goes back to the same idea of the

“Harberger triangle”; that is using statistics, typically elasticities, that can be estimated non-parametrically, and one can approximate various quantities, such as dead-weight loss, etc., that are relevant for the design and the evaluation of public policies. However, as far as optimal taxation is concerned, in general the idea only works for the computation of local solutions (for instance the top marginal tax rate). An interesting special case presented by Saez (2001) is the computation of the optimal top marginal tax rate, above income level  $\tilde{z}$ . Assuming quasi-linear preferences and constant elasticity  $e$ , it turns out that

$$\frac{\tau}{1-\tau} = \frac{[1-\tilde{g}] \left[ 1 - \frac{\tilde{z}}{z^m} \right]}{e}$$

where  $z^m$  is the average income of households with  $z \geq \tilde{z}$  and  $\tilde{g}$  is the social weight attached to those same households. Empirical one finds that, for sufficiently high  $\tilde{z}$ , the ratio  $\tilde{z}/z^m$  is approximately constant. Therefore, the top marginal tax rate can be directly computed as a function of the elasticity  $e$  and of the social preferences summarized by  $\tilde{g}$ .

The approach pioneered by Mirrlees (1971) and innovated by Saez (2001, 2002) is a fundamental theoretical framework for addressing the design of optimal tax-transfer mechanisms. However, so far, its empirical applications suffer from three main shortcomings due to the assumptions made in order to get practical analytical solutions. First, Mirrlees (1971) and Saez (2001), among others, only cover interior solutions and therefore only intensive labour supply responses are considered. Saez (2002) presents a discrete choice model that includes extensive responses but introduces special restrictive assumptions on the intensive responses.<sup>12</sup> Second, the empirical implementations of the analytical approach so far have considered individuals, not couples.<sup>13</sup> Third, most empirical applications assume quasi-linear preferences, no income effects and fixed labour supply elasticities.

To overcome the shortcomings of the simulation exercises based on theoretical optimal taxation results, recent contributions have proposed an alternative, or complementary, computational approach (Aaberge & Colombino, 2012, 2013b; Blundell & Shephard, 2012; Ericson & Flood, 2012; Islam & Colombino, 2018). Modern microeconomic models of labour supply can accommodate many realistic features such as simultaneous decisions of household members, non-unitary mechanisms of household decisions, decisions at both the intensive and extensive margins, complicated constraints and opportunity sets, multidimensional heterogeneity of both households and jobs, quantitative constraints etc. It is simply not feasible, at least so far, to obtain analytical

solutions for the optimal taxation problem in such environments. The computational approach combines microeconomic modelling, microsimulation and numerical optimization. The microeconomic model, which simulates the agents' choices by utility maximization, is embedded into a global maximization algorithm that solves the social planner's problem, that is, the maximization of a social welfare function subject to the public budget constraint.

The method, as presented in Aaberge and Colombino (2013), can be formulated as in Equation 16 below. Household  $n$  can choose a "job" within an opportunity set  $B_n$ . Each job is defined by a vector of wage rates  $w$ , a vector of hours of work  $h$  and other characteristics  $j$  (unobserved by the analyst). Given gross earnings  $w'h$  and gross unearned income  $I$ , net available income is determined by a tax-transfer function  $c = f(w'h, I; \mathcal{G})$  defined up to a vector of parameters  $\mathcal{G}$ . For any given tax-transfer rule (that is, any given value of  $\mathcal{G}$ ) the choices by the households are simulated by running a microeconomic model that allows for a very flexible representation of heterogeneous preferences and opportunity sets, it covers both singles and couples, accounts for quantity constraints and is able to treat any tax-transfer rule however complex. Note that it would be hopeless to look for analytical solutions of an optimal taxation problem in such an environment. The choices made by the  $N$  agents result in  $N$  positions  $(c_1, h_1, j_1), (c_2, h_2, j_2), \dots, (c_N, h_N, j_N)$ , which are then evaluated by the social planner according to a social welfare function  $\mathcal{W}$ . The Social Planner's problem therefore consists of searching for the value of the parameters  $\mathcal{G}$  that maximizes  $\mathcal{W}$  subject to the following constraints: (i) the various positions  $(c_1, h_1, j_1), \dots, (c_N, h_N, j_N)$  result from utility-maximizing choices on the part of the households (incentive-compatibility constraints); (ii) the total net tax revenue must attain a given amount  $R$  (public budget constraint). The optimal taxation problem

$$\begin{aligned} & \max_{\mathcal{G}} \mathcal{W} \left( U_1(c_1, h_1, j_1), U_2(c_2, h_2, j_2), \dots, U_N(c_N, h_N, j_N) \right) \\ & \text{s.t.} \\ & (c_n, h_n, j_n) = \arg \max_{(w, h, j) \in B_n} U_n(c, h, j) \text{ s.t. } c = f(wh, I_n; \mathcal{G}), \forall n \\ & \sum_{n=1}^N (w_n h_n + I_n - f(w_n h_n, I_n; \mathcal{G})) \geq R. \end{aligned} \quad (16)$$

is solved computationally by iteratively simulating the household choices for different values of  $\mathcal{G}$  until  $\mathcal{W}$  is maximized. As indicated above, several recent contributions identify optimal tax-benefit rules by employing random utility models of labour supply together with microsimulation and, some version of, the social evaluation framework presented in Section 5.5 below. Aaberge and

Colombino (2013) identify optimal income tax regimes in Norway within a 10-parameter family of piecewise linear systems based on rank-dependent social welfare functions with different inequality aversion profiles. A similar exercise for Italy, where the adopted social welfare criteria account for inequality-of-opportunity, has been considered by Aaberge and Colombino (2012). Blundell and Shephard (2012) have designed an optimal tax-benefit rule for low-income families with children in the UK. Colombino and Narazani (2013) and Colombino (2015) have focussed on alternative basic income-support in Italy. Islam and Colombino (2018) have identified optimal tax-transfer rules in the NIT+FT class for a sample of European countries. As opposed to the theory-based optimal tax exercises, the micro-econometric simulation approach allows for a much more flexible representation of households' preferences and choice opportunities and permits analysis of more complicated tax-benefit rules. This has significant implications upon the results. For example, Aaberge and Colombino (2013), for each of four different social welfare functions with inequality aversion profiles that range from neutrality to strong downside inequality aversion, identify the tax system that maximizes social welfare within a class of 10 parameter tax rules. The results show that the marginal tax rates of each of the optimal tax systems turned out to be monotonically increasing with income and that more egalitarian social welfare functions tended to imply more progressive tax rules. Moreover, the optimal bottom marginal tax rate is negative, suggesting a mechanism close to policies like the Working Families Tax Credit in the UK, the Earned Income Tax Credit in the USA and the In-Work Tax Credit in Sweden. The overall emerging picture is somehow close to Saez (2002) and Diamond and Saez (2011) but is in sharp contrast with most of the results obtained by the numerical exercises based on Mirrlees (1971) or Saez (2001). The typical outcomes of the latter exercises envisage a positive lump-sum transfer which is progressively taxed away by very high marginal tax rates on lower incomes, in combination with a proportional or slightly increasing tax rate on higher incomes. Islam and Colombino (2018) show a large heterogeneity of results across different countries and —within the NIT+FT class— find that most of the optimal rules present a concave NIT profile, that is, the phasing-out marginal rate applied to subsidised incomes is lower than the (flat) tax rate applied to higher incomes. Overall, the results obtained with the microsimulation approach seems to support what suggested by Tuomala (2010): the theory-based results might be enforced by the restrictive assumptions made on the preferences, the elasticities and the distribution of productivities (or wage rates), which in turn might conflict with the empirical evidence provided by microeconomic labour supply studies.

## 5.5 Social evaluation of policy reforms

### 5.5.1 *Individual welfare functions*

As explained above, empirical microeconomic models of labour supply are helpful tools for simulating the effects on households' labour supply and income from changes in tax and benefit systems or from changes in distributions of wage rates and hours of work offered by the demand side of the labour market. It is straightforward to provide a summary of changes in employment rates and distributions of hours of work and income. However, a social planner needs information that makes it possible to compare individuals' level of welfare before and after a policy change and thus who is gaining and who is losing on the policy change. It is, however, not obvious how one should make a social evaluation of the policy effects when the individuals' welfare is a function of income and leisure. The estimated utility functions, or their systematic parts, might emerge as a useful basis for making social evaluations of welfare. However, since the behaviour of an individual is invariant with respect to monotonic transformations of the utility function, we face two problems. The first one concerns the construction of specific cardinal utility functions to represent the consumption/leisure preferences of individuals/households, and the second concerns the lack of convincing justification for comparing arbitrarily chosen individual cardinal utility functions and use them as arguments in a social welfare function (see for instance the thorough discussion provided by Hammond, 1991). The origin of the problem is as stated by Hume that one cannot derive an "ought" from an "is". The common practice of basing social evaluations on distributions of individual-specific money metric measures of utility differences like equivalent and compensating variation disregards the interpersonal comparability problem, which makes it difficult to judge the ethical significance of this approach. To circumvent these problems, Deaton and Muellbauer (1980) and Hammond (1991) propose to use a common utility function as a tool for making interpersonal comparisons of welfare, since it by definition contains interpersonal comparability of both welfare levels and welfare differences. The common utility function is supposed to capture the preferences of the social planner, whereas the individual/household-specific utility functions solely are assumed to capture the consumption/leisure preferences of individuals/households. The latter can be used to simulate the behaviour of individuals/households under alternative tax/benefit systems, whereas the former is designed to be used for evaluating the outcomes of simulation exercises. As argued by Aaberge and Colombino (2013) a plausible approach is to assume that the social planner exploits the information provided by the consumption/leisure choices of the individuals/households (and moreover accounts for large heterogeneity in the availability of different jobs in the market) by estimating the common



utility function. Below we will provide an explanation of the specific version of the common utility approach employed by Aaberge and Colombino (2013) for designing optimal taxes based on a microeconomic model of labour supply. Since households differ regarding size and composition, it is required to construct a common utility function that justifies comparisons of individual welfare for individuals. The common utility function, individual welfare function,  $V$  is to be interpreted just as the input of a social welfare function and thus differs from the role played by the actual utility function  $U$  for households. The individual welfare function ( $V$ ) is assumed to have a functional form that is identical to the basic functional form of the systematic part of the positive utility function  $U$ , which means that the heterogeneity of the parameters of  $U$  has been removed. Thus,  $V$  is defined by (Equation 17)

$$V(y, h) = \gamma_2 \left( \frac{y^{\gamma_1} - 1}{\gamma_1} \right) + \gamma_4 \left( \frac{L^{\gamma_3} - 1}{\gamma_3} \right) \quad (17)$$

where  $L = 1 - (h/8736)$ , and  $y$  is the individual's income after tax defined by (Equation 18):

$$y = \begin{cases} c = f(wh, I) & \text{for singles} \\ \frac{c}{\sqrt{2}} = \frac{1}{\sqrt{2}} f(w_F h_F, w_M h_M, I) & \text{for married/cohab. individuals.} \end{cases} \quad (18)$$

Thus, couples' incomes are transformed into comparable individual-specific incomes by dividing the couple incomes by the square root of 2. The parameters of  $V(.,.)$  are estimated with Equation 14, where  $v$  is replaced by  $V$ .

Alternative and promising approaches aiming at respecting individual (consumption/leisure) preferences in welfare analyses have been proposed by Piacquadio (2017), Fleurbaey (2008), and Fleurbaey and Maniquet (2006). The approach discussed in the two latter papers has been applied in analyses of labour supply (Bargain et al., 2013; Decoster & Haan, 2015). However, as acknowledged by Decoster and Haan (2015), the choice of a specific preference respecting welfare metric might have a significant impact on the result of the welfare evaluation, and moreover shows to depend on the degree of emphasis the welfare metric places on willingness-to-work. Thus, depending on the chosen metric, a work averse or work loving individual will be favoured, which means that the social planner faces the problem of giving more or less weight to people with preferences that exhibit low or high willingness-to-work.

King (1983) proposes an approach where different preferences are represented by different characteristics or parameters  $Z_i$  within a common parametric utility function. The characteristics account for a different productivity in obtaining utility from the opportunities available in the budget set. Let  $V^*(w_i, I_i, Z_i)$  be the maximum utility attained by household  $i$  given the budget define by  $(w_i, I_i)$ . We consider reference characteristics  $Z_R$  and a reference budget  $(w_R, I_R)$  and the corresponding maximum utility  $V^*(w_R, I_R, Z_R)$ . The comparable money-metric index  $\omega_i$  is then defined by (Equation 19):

$$V^*(w_R, \omega_i, Z_R) = V^*(w_i, I_i, Z_i) \quad (19)$$

Empirical applications of this approach are provided by King (1983), Aaberge et al. (2004), and Islam and Colombino (2018).

A different way to circumvent the interpersonal comparability problem consists in avoiding interpersonal comparisons altogether and basing the social evaluation exclusively on intrapersonal comparisons of utility levels, which of course is less informative. A proper application of the ordinal criterion would require defining the optimal tax in a different way, for example the rule that maximizes the number of winners. However, since the winners might be the individuals with the highest pre-reform welfare levels, the ordinal criterion does obviously not account for distributional effects and may for that reason be considered as an inappropriate social evaluation approach.

### 5.5.2 *Social welfare functions – the primal and dual approach*

The informational structure of the individual welfare functions (defined by the common utility Equation 17 or Piacquadio's and Fleurbaey's preference respecting welfare metrics) allows comparison of welfare levels as well as gains and losses of different individuals due to a policy change. Comparison of distributions of individual welfare, induced for example by alternative hypothetical tax reforms, might be made in terms of dominance criteria of first- and second degree. However, since distribution functions normally intersect even second-degree dominance may not provide an unambiguous ranking of the distributions in question, but it would in any case be helpful to quantify social welfare by applying either a primal or a dual social welfare function.

The “primal approach” is analogue to the inequality framework developed by Atkinson (1970), while the “dual approach” is analogue to the rank-dependent measurement of inequality introduced

by Weymark (1981) and Yaari (1988). As is well known the Independence Axiom justifies the following family of social welfare functions (Equation 20),

$$W(F) = \int_0^{\infty} u(x) dF(x) \quad (20)$$

where  $F$  is a distribution with mean  $\mu$  of the individual welfare  $V$ , and  $u$  is a non-decreasing concave evaluation function of individual welfare levels that reflects the preferences of a social planner who support the Independence Axiom. As demonstrated by Atkinson (1970),  $W$  can be represented by the equally distributed equivalent welfare level defined by Equation 21:

$$\xi(F) = u^{-1}(W(F)) \quad (21)$$

Thus,  $\xi(F)$  is the equally distributed individual welfare level that would yield the same level of social welfare as the actual distribution  $F$ . Since  $\xi(F) \leq \mu$ , Atkinson (1970) used  $\xi(F)$  as a basis for defining the following family of inequality measures (Equation 22):

$$I(F) = 1 - \frac{\xi(F)}{\mu} \quad (22)$$

The following specific family of social welfare functions and associated inequality measures were introduced by Atkinson (1970) (Equation 23),

$$\xi(F) = \left( \int_0^{\infty} x^{1-\theta} dF(x) \right)^{\frac{1}{1-\theta}} \quad (23)$$

where  $\theta \geq 0$  defines the degree of inequality aversion of the social welfare function.

A similar structure is captured by the family of rank-dependent welfare functions (Weymark, 1981; Yaari, 1988) (Equation 24):

$$W_k^*(F) = \int_0^1 p_k(t) F^{-1}(t) dt \quad (24)$$

where  $F^{-1}$  is the left inverse of the cumulative distribution function of the individual welfare levels  $V$  with mean  $\mu$ , and  $p_k(t)$  —a positive concave weight-function defined on the unit interval— represents the preferences of the social planner and depends on an inequality-aversion parameter  $k$ .<sup>14</sup> The social welfare functions (Equation 24) can be given a similar normative justification as for the family (Equation 20). We refer to Aaberge and Colombino (2014) for the specification of the

weight function  $p_k(t)$ . As suggested by Weymark (1981) and Aaberge (2007), the index (Equation 25) can be used as a measure of inequality:

$$C_k = 1 - \frac{W_k^*}{\mu}, \quad i = 1, 2, \dots \quad (25)$$

The inequality indices (Equations 22 and 24) are invariant with respect to multiplicative constants. Alternatively, one might define indices that are invariant with respect to additive constants. An example is provided by Kolm (1976), where the index of inequality is:

$$K = \frac{1}{\alpha} \ln \int \exp\{-\alpha(x - \mu)\} dF(x) \quad (26)$$

where  $\alpha > 0$  is a parameter that exhibits inequality aversion. The corresponding index of social welfare can be defined as  $W = \mu - K$ . This approach is adopted by Islam and Colombino (2018). A similar index is also used by Blundell and Shephard (2012). Apart from the different theoretical assumptions, there might be practical issues that drive the preference among the different indices. For example, in empirical applications a rescaling of the arguments of the social welfare indices is often required or convenient. Depending on the different circumstances, a multiplicative or rather an additive rescaling might turn out as more appropriate.

## 6 CONCLUSIONS AND FUTURE PERSPECTIVES

The original concept of microsimulation envisaged large models of the entire economic (or even socio-economic) system —as an alternative to the then dominating large macroeconomic models— including behavioural responses. The events took a different route. On the one hand, the first successful implementations of microsimulation models at the policy level were non-behavioural. On the other hand, the researchers working on microeconomic models of labour supply started using microsimulation tools for policy design and evaluation. In this paper, we have illustrated the current labour supply modelling strategies and their possible evolutions, together with their policy applications that use microsimulation methods. Further developments, both on the microsimulation algorithms side and on the microeconomic side, might or might not favour a development of a stronger link between large microsimulation algorithms and behavioural labour supply analysis. The general problem is that there is a trade-off between the increasing theoretical sophistication of labour supply models (that is, stochastic dynamic programming models, intra-household allocation or collective models, etc.) and their flexibility in interacting with other models

representing different segments of the economic system. However, the approach currently adopted in most of the labour supply modelling literature, that is the RUM/RURO approach, at the moment represents an excellent compromise between increasing sophistication and tractability/interactions within larger simulation projects. Addressing more complex tax-transfer policies, adding other dimension of choice (besides hours of work) or introducing dynamics and intertemporal choices, do not change the basic logical and computational structure of RUM/RURO models. Their typical discrete representation of the opportunity sets is naturally matched to the logic of discrete states and discrete choices prevailing in microsimulation since its origins. Furthermore, microsimulation provides an ideal platform for addressing issues that are hard, if not impossible, to tackle analytically, that is identifying optimal tax-transfer policies, comparing alternative theory of choice (for instance utility maximization vs. agent-based models) or exploring the implications of alternative social welfare evaluation criteria.

## REFERENCES

- Aaberge, R. (2007). Gini's nuclear family. *Journal of Economic Inequality*, 5(3), 305–322.
- Aaberge, R., & Colombino, U. (2009). *Accounting for Family Background when Designing Optimal Income Taxes: A Microeconomic Simulation Analysis*.
- Aaberge, R., & Colombino, U. (2012). Accounting for family background when designing optimal income taxes: a microeconomic simulation analysis. *Journal of Population Economics*, 25(2), 741–761.
- Aaberge, R., & Colombino, U. (2013). Using a Microeconomic Model of Household Labour Supply to Design Optimal Income Taxes. *Scandinavian Journal of Economics*, 115(2), 449–475.
- Aaberge, R., & Colombino, U. (2014). Labour Supply Models. In C. O'Donoghue (Ed.), *Handbook in Microsimulation Modelling – Contributions to Economic Analysis* (pp. 167–221). UK: Emerald Group Publishing Limited.
- 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., & Strøm, S. (1999). Labour supply in Italy: An empirical analysis of joint household decisions, with taxes and quantity constraints. *Journal of Applied Econometrics*, 14(4), 403–422.
- Aaberge, R., Colombino, U., & Strøm, S. (2004). Do more equal slices shrink the cake? An empirical

- investigation of tax-transfer reform proposals in Italy. *Journal of Population Economics*, 17(4), 767–785.
- Aaberge, R., Colombino, U., & Wennemo, T. (2009). Evaluating alternative representations of the choice sets in models of labor supply. *Journal of Economic Surveys*, 23(3), 586–612.
- Aaberge, R., Dagsvik, J. K., & Strøm, S. (1995). Labor Supply Responses and Welfare Effects of Tax Reforms. *The Scandinavian Journal of Economics*, 97(4), 635–659.
- Aaberge, R., & Flood, L. (2013). *U.S. versus Sweden: The Effect of Alternative In-Work Tax Credit Policies on Labour Supply of Single Mothers* (IZA Discussion Paper No. 7706).
- Arrufat, J. L., & Zabalza, A. (1986). Female Labor Supply with Taxation, Random Preferences, and Optimization Errors. *Econometrica*, 54(1), 47–63.
- Atkinson, A. B. (1970). On the Measurement Of Inequality. *Journal of Econometric Theory*, 2, 244–263.
- Atkinson, A. B. (1996). *Public Economics in Action: The Basic Income/Flat Tax Proposal*. Oxford University Press.
- Atkinson, A. B. (2015). *Inequality. What Can be Done?* Harvard University Press.
- Bargain, O., Decoster, A., Dolls, M., Neumann, D., Peichl, A., & Siegloch, S. (2013). Welfare, labor supply and heterogeneous preferences: Evidence for Europe and the US. *Social Choice and Welfare*, 41(4), 789–817.
- Bargain, O., & Doorley, K. (2017). The Effect of Social Benefits on Youth Employment: Combining Regression Discontinuity and a Behavioral Model. *Journal of Human Resources*, 52(4), 1032–1059.
- Bargain, O., Orsini, K., & Peichl, A. (2014). Comparing Labor Supply Elasticities in Europe and the United States. *Journal of Human Resources*, 49(3), 723–838.
- Bargain, O., & Peichl, A. (2016). Own-wage labor supply elasticities: variation across time and estimation methods. *IZA Journal of Labor Economics*.
- Ben-Akiva, M., & Lerman, S. (1985). *Discrete choice analysis: theory and application to travel demand*. MIT Press, Cambridge, MA.
- Ben-Akiva, M., & Watanatada, T. (1981). Application of a Continuous Spatial Choice Logit Model. In C. F. Manski & D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press.
- Bhuller, M., Vestad, O., & Aaberge, R. (2016). *The effects of tax credit on immigrants labour supply (in Norwegian)* (No. 38). Statistics Norway.
- Bloemen, H. G. (2010). An empirical model of collective household labour supply with non-participation. *Economic Journal*, 120(543), 183–214.

- Blomquist, S., & Newey, W. (2002). Nonparametric estimation with nonlinear budget sets. *Econometrica*, 70(6), 2455–2480.
- Blundell, R. (2012). Tax policy reform: The role of empirical evidence. *Journal of the European Economic Association*, 10(1), 43–77.
- Blundell, R., Chiappori, P. A., Magnac, T., & Meghir, C. (2007). Collective labour supply: Heterogeneity and non-participation. *Review of Economic Studies*, 74(2), 417–445.
- Blundell, R., & Shephard, A. (2012). Employment, hours of work and the optimal taxation of low-income families. *Review of Economic Studies*, 79(2), 481–510.
- Bosmans, K., Decancq, K., & Ooghe, E. (2015). What do normative indices of multidimensional inequality really measure? *Journal of Public Economics*, 130, 94–104.
- Bourguignon, F., & Spadaro, A. (2006). Microsimulation as a tool for evaluating redistribution policies. *Journal of Economic Inequality*, 4(1), 77–106.
- Brewer, M., Saez, E., & Shephard, A. (2008). *Means-testing and tax rates on earnings: Prepared for the Report of a Commission on Reforming the Tax System for the 21st Century, Chaired by Sir James Mirrlees*. Oxford University Press.
- Chetty, R. (2009). Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods. *Annual Review of Economics*, 1(1), 451–488.
- Chetty, R. (2015). Behavioral Economics and Public Policy: A Pragmatic Perspective. *American Economic Review*, 105(5), 1–33.
- Chiappori, P.-A. (1988). Rational Household Labor Supply. *Econometrica*, 56(1), 63–90.
- Clavet, N., Duclos, J., & Lacroix, G. (2013). *Fighting Poverty: Assessing the Effect of Guaranteed Minimum Income Proposals in Québec* (IZA No. 7283).
- Colombino, U. (2013). A New Equilibrium Simulation Procedure with Discrete Choice Models. *International Journal of Microsimulation*, 6(3), 25–49.
- Colombino, U. (2015). Five Crossroads on the Way to Basic Income. An Italian Tour. *Italian Economic Journal*, 1(3), 353–389.
- Colombino, U., & Narazani, E. (2013). Designing a Universal Income Support Mechanism for Italy: An Exploratory Tour. *Basic Income Studies*, 8(1), 1–17.
- Creedy, J., & Duncan, A. (2002). Behavioural Microsimulation with Labour Supply Responses. *Journal of Economic Surveys*, 16(1), 1–39.
- Creedy, J., & Duncan, A. (2005). Aggregating labour supply and feedback effects in microsimulation. *Australian Journal of Labour Economics*, 8(3), 277–290.
- Dagsvik, J. K. (1994). Discrete and Continuous Choice, Max-Stable Processes, and Independence from Irrelevant Attributes. *Econometrica*, 62(5), 1179.

- Dagsvik, J. K., Locatelli, M., & Strøm, S. (2009). Tax reform, sector-specific labor supply and welfare effects. *Scandinavian Journal of Economics*, 111(2), 299–321.
- Deaton, A., & Muellbauer, J. (1980). *Economics and Consumer Behavior*. Cambridge: University Press.
- Decoster, A. M. J., & Haan, P. (2015). Empirical welfare analysis with preference heterogeneity. *International Tax and Public Finance*, 22(2), 224–251.
- Del Boca, D., & Flinn, C. (2012). Endogenous household interaction. *Journal of Econometrics*, 166(1), 49–65.
- Diamond, P., & Saez, E. (2011). The case for a progressive tax: From basic research to policy recommendations. *Journal of Economic Perspectives*, 25(4), 165–190.
- Domencich, T. A., & McFadden, D. L. (1975). Urban travel demand: A behavioral analysis. *North-Holland Publishing Co.* Reprinted by The Blackstone Company: Mount Pleasant, MI, 1996.
- Ericson, P., & Flood, L. (2012). A Microsimulation Approach to an Optimal Swedish Income Tax. *International Journal of Microsimulation*, 5(2), 2–21.
- Feldstein, M. (1995). The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act. *Journal of Political Economy*, 103(3), 551–572.
- Fleurbaey, M. (2008). *Fairness, responsibility, and welfare*. Oxford: University Press.
- Fleurbaey, M., & Maniquet, F. (2006). Fair income tax. *Review of Economic Studies*, 73(1), 55–83.
- Friedman, M. (1962). *Capitalism and freedom*. University of Chicago Press.
- Fuest, C., Peichl, A., & Schaefer, T. (2008). Is a flat tax reform feasible in a grown-up democracy of Western Europe? A simulation study for Germany. *International Tax and Public Finance*, 15(5), 620–636.
- Gong, X., & van Soest, A. (2002). Family Structure and Female Labor Supply in Mexico City. *The Journal of Human Resources*, 37(1), 163.
- Haan, P., & Wrohlich, K. (2011). Can child care policy encourage employment and fertility? Evidence from a structural model. *Labour Economics*, 18(4), 498–512.
- Hammond, P. (1991). Interpersonal comparisons of utility: Why and how they are and should be made. *Interpersonal Comparisons of Well-Being*, 200–254.
- Hausman, J. A. (1979). The econometrics of labor supply on convex budget sets. *Economics Letters*, 3(2), 171–174.
- Hausman, J. A. (1985). Taxes and labor supply. In A. J. Auerbach & M. Feldstein (Eds.), *Handbook of Public Economics* (pp. 213–263). Elsevier.
- Heckman, J. J. (1974). Effects of Child-Care Programs on Women's Work Effort. *Journal of Political Economy*, 82(2), 136–163.
- Horstschräer, J., Clauss, M., & Schnabel, R. (2010). *An Unconditional Basic Income in the Family Context:*



- Labor Supply and Distributional Effects* (ZEW Center for European Economic Research No. 10–091).
- Hoynes, H. W. (1996). Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation Under AFDC-UP. *Econometrica*, 64(2), 295–332.
- Islam, N., & Colombino, U. (2018). *The case for negative income tax with flat tax in Europe: an empirical optimal taxation exercise* (Society for the Study of Economic Inequality (ECINE) No. 454).
- Jacquet, L., Lehmann, E., & Van der Linden, B. (2013). Optimal redistributive taxation with both extensive and intensive responses. *Journal of Economic Theory*, 148(5), 1770–1805.
- Jessen, R., Rostam-Afschar, D., & Steiner, V. (2017). Getting the Poor to Work: Three Welfare-Increasing Reforms for a Busy Germany. *FinanzArchiv: Public Finance Analysis*, 73(1), 1–41.
- Kalb, G. R. (2000). *Labour Supply and Welfare Participation in Australian Two-Adult Households: Accounting for Involuntary Unemployment and the “Cost” of Part-time Work*. Victoria University, Centre of Policy Studies/IMPACT Centre.
- Keane, M., & Moffitt, R. (1998). A Structural Model of Multiple Welfare Program Participation and Labor Supply. *International Economic Review*, 39(3), 553–589.
- Keane, M. P. (2010). A Structural Perspective on the Experimentalist School. *Journal of Economic Perspectives*, 24(2), 47–58.
- Keane, M. P. (2011). Labor Supply and Taxes: A Survey. *Journal of Economic Literature*, 49(4), 961–1075.
- Keane, M. P., & Wolpin, K. I. (2002a). Estimating Welfare Effects Consistent with Forward-Looking Behavior. Part I: Lessons from a Simulation Exercise. *Journal of Human Resources*, 37(3), 570–599.
- Keane, M. P., & Wolpin, K. I. (2002b). Estimating Welfare Effects Consistent with Forward-Looking Behavior. Part II: Empirical Results. *The Journal of Human Resources*, 37(3), 600.
- King, M. A. (1983). Welfare analysis of tax reforms using household data. *Journal of Public Economics*, 21(2), 183–214.
- Kleven, J. H., Kreiner, C. T., & Saez, E. (2009). The Optimal Income Taxation of Couples. *Econometrica*, 77(2), 537–560.
- Klevmarken, N. A. (1997). *Modelling Behavioural Response in EUROMOD*. Faculty of Economics, University of Cambridge.
- Kolm, S.-C. (1976). Unequal Inequalities. *Journal of Econometric Theory*, 12, 416–442.
- Kornstad, T., & Thoresen, T. O. (2007). A discrete choice model for labor supply and childcare. *Journal of Population Economics*, 20(4), 781–803.
- Li, J., & O’Donoghue, C. (2013). A survey of dynamic microsimulation models: uses, model

- structure and methodology. *International Journal of Microsimulation*, 6, 3–55.
- Lucas, R. E. (1976). Econometric policy evaluation: a critique. *Carnegie-Rochester Conference Series on Public Policy*, 1(1), 19–46.
- Marschak, J. (1953). Economic measurements for policy and prediction. In W. Hood & T. Koopmans (Eds.), *Studies in Econometric Method* (pp. 1–26). New York: Wiley.
- Matzkin, R. L. (2013). Nonparametric Identification in Structural Economic Models. *Annual Review of Economics*, 5(1), 457–486.
- McElroy, M. B., & Horney, M. J. (1981). Nash-Bargained Household Decisions: Toward a Generalization of the Theory of Demand. *International Economic Review*, 22(2), 333–349.
- McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, 3(4), 303–328.
- McFadden, D. L. (1984). Econometric analysis of qualitative response models. In Z. Griliches & M. Intriligator (Eds.), *Handbook of Econometrics* (pp. 1395–1457). Elsevier.
- McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447–470.
- Mirrlees, J. A. (1971). An Exploration in the Theory of Optimal Taxation. *Review of Economic Studies*, 38(2), 175–208.
- Moffitt, R. A. (2003). The Negative Income Tax and the Evolution of U.S. Welfare Policy. *Journal of Economic Perspectives*, 17(3), 119–140.
- Neugart, M., & Richiardi, M. G. (2012). *Agent-based models of the labor market* (Revelli Working Papers Series No. 125). LABORatorio R.
- Orcutt, G. H. (1957). A New Type of Socio-Economic System. *The Review of Economics and Statistics*, 39(2), 116.
- Paulus, A., & Peichl, A. (2008). *Effects of flat tax reforms in Western Europe on equity and efficiency* (FiFo-CPE Discussion Paper No. 8–4).
- Peichl, A. (2014). Flat-rate tax systems and their effect on labor markets. *IZA World of Labor*, 2014–61.
- Peichl, A., & Siegloch, S. (2012). Accounting for labor demand effects in structural labor supply models. *Labour Economics*, 19(1), 129–138.
- Piacquadio, P. G. (2017). A fairness justification of utilitarianism. *Econometrica*, 85(4), 1261–1276.
- Saez, E. (2001). Using elasticities to derive optimal income tax rates. *Review of Economic Studies*, 68(1), 205–229.
- Saez, E. (2002). Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses. *The Quarterly Journal of Economics*, 117(3), 1039–1073.

- Scutella, R. (2004). *Moves to a Basic Income-Flat Tax System in Australia: Implications for the Distribution of Income and Supply of Labour* (Melbourne Institute of Applied Economic and Social Research No. 5).
- Sommer, M. (2016). *A Feasible Basic Income Scheme for Germany: Effects on Labor Supply, Poverty, and Income Inequality*. Springer.
- Swann, C. A. (2005). Welfare Reform When Recipients Are Forward-Looking. *The Journal of Human Resources*, 40(1), 31–56.
- Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press. Cambridge University Press.
- Tuomala, M. (2010). On optimal non-linear income taxation: numerical results revisited. *International Tax and Public Finance*, 17(3), 259–270.
- Van Parijs, P. (1995). *Real Freedom for All*. Oxford University Press.
- Van Soest, A. (1995). Structural Models of Family Labor Supply: A Discrete Choice Approach. *The Journal of Human Resources*, 30(1), 63–88.
- Weymark, J. A. (1981). Generalized Gini inequality indices. *Mathematical Social Sciences*, 1(4), 409–430.
- Wolpin, K. I. (1996). Public-Policy Uses of Discrete-Choice Dynamic Programming Models. *American Economic Review*, 86(2), 427–432.
- Yaari, M. E. (1988). A controversial proposal concerning inequality measurement. *Journal of Economic Theory*, 44(2), 381–397.
- Zabalza, A., Pissarides, C., & Barton, M. (1980). Social security and the choice between full-time work, part-time work and retirement. *Journal of Public Economics*, 14(2), 245–276.

## NOTES

<sup>1</sup> Up to that period, the typical procedure consisted of evaluating elasticities or policy effects with reference to the “average” or in some sense “representative” household. Even the path-breaking contributions to structural labour supply modelling (see Heckman (1974) or Hausman (1985) adopted the “average household” approach or computed behavioural responses for different “types” of households.

<sup>2</sup> See Aaberge et al. (2009) and Aaberge and Colombino (2014) for a comparison between RUM and previous approaches to modelling labour supply.

<sup>3</sup> A motivation for preferring this interpretation of  $\epsilon$  in DC models is the relatively small number of values of  $b$  that are typically allowed to belong to the opportunity set, in many cases just three (the midpoints of three hour brackets). Since the observed distribution of hours worked is much more dispersed, it might make sense to allow for a measurement/optimization error.

<sup>4</sup> The derivation of the Conditional Logit expression for utility maximisation under the assumption that the utility random components are i.i.d. Type I extreme value distributed is due to McFadden (1974). It is conventional to call Conditional Logit a Multinomial Logit model with generic attributes (that is attributes, like hours or income, whose values vary across alternatives).

<sup>5</sup> Examples of applications to labour supply are provided by Neugart and Richiardi (2012).

<sup>6</sup> Chetty (2015; 2009) provides useful discussions of the links between structural models, partial identification of structural parameters and departures from utility maximization.

<sup>7</sup> The systematic analysis of the statistical properties of alternative methods for producing predictions is more advanced in other areas where RUM models are used, for instance Ben-Akiva and Lerman (1985).

<sup>8</sup> A more extensive survey of policy applications can be found in Aaberge and Colombino (2014). The role of empirical evidence in view of the design of tax-benefit reforms is surveyed by Blundell (2012).

<sup>9</sup> Moffit (2003) provides an analysis of the NIT proposal and of the policy debate around it.

<sup>10</sup> See Islam and Colombino (2018) for an interpretation of NIT as a general class that include CBI and UBI as a special limit cases.

<sup>11</sup> if  $z = (1-T)wh$ , then  $e(z)$  is also equal to  $(1 + \text{elasticity of } h \text{ with respect to } w)$ .

<sup>12</sup> Jacquet, Lehmann, and Van der Linden (2013) present a different theoretical model with different implications.

<sup>13</sup> A recent theoretical contribution is Kleven, Kreiner, and Saez (2009).

<sup>14</sup> Note, the Equations 17 and 20 (or Equation 24) can be considered as two-stage approaches for measuring social welfare where the first stage consists of using the common utility function to aggregate the two goods (consumption and leisure) for each individual into a measure of well-being and the second stage to aggregate the well-being across individuals into a measure of social welfare. As demonstrated by Bosmans, Decancq, and Ooghe (2015) the two-stage approach can be given an axiomatic normative justification.