rivista di statistica ufficiale

n. 2
2015

Temi trattati

Istat
Istituto Nazionale di Statistica

rivista di statistica ufficiale

n. 2
2015

Temi trattati

Istat

Per contattare la redazione o per inviare lavori scrivere a:
Segreteria del Comitato di redazione della Rivista di Statistica Ufficiale
All'attenzione di Gilda Sonetti
Istat – Via Cesare Balbo, 16 – 00184 Roma
e-mail: rivista@istat.it

rivista di statistica ufficiale

n. 2/2015

Periodico quadrimestrale
ISSN 1828-1982

Registrato presso il Tribunale di Roma
n. 339 del 19 luglio 2007

Istituto nazionale di statistica
Via Cesare Balbo, 16 – Roma

Stampato nel mese di Settembre 2015
da System Graphic srl
Via di Torre S. Anastasia, 61 - Roma

FSC www.fsc.org
MISTO
Carta da fonti gestite in maniera responsabile
FSC® C117539

The Istat Microsimulation Models[1]

Ugo Colombino[2]

Sommario

Obiettivo di questo lavoro è di inserire il modello di microsimulazione sulle famiglie sviluppato dall'Istat all'interno dell'evoluzione dei modelli statici e comportamentali. Oltre a una panoramica internazionale sui modelli si descrivono le implicazione dei modelli statici per l'analisi delle policy.

Parole chiave: microsimulazione, tassazione.

Abstract

The aim of this paper is to relate the new microeconometric model on households developed by Istat to the development of the static and behavioural models presented in the literature. Both a survey on the international experiences and a focus on the implication of the static model for policy evaluation are presented.

Keywords: Microsimulation, taxation.

1. Introduction

The Istat new microsimulation models described in this volume come out at a moment of maturity of microsimulation research, when the respective roles of static and behavioural models - and their relationships and interactions – have been made clear and productive, after decades of encounters, conflicts and re-encounters. It is instructive to summarize the process that brought us where we stand now (Section 2). Then we will look at the current state-of-the-art in static modelling in Italy and elsewhere (Section 3). In Section 4, we address the issue of how to interpret the static microsimulation results from the policy point-of-view. We also suggest some procedures that have the potential of enriching the static models with elements of behavioural response without having to develop a fully specified structural behavioural model. Section 5 contains the conclusions.

[1] The opinions expressed are those of the author and do not entail the responsibility of Istat.

[2] Professor of Economics, Dipartimento di Economia e Statistica Cognetti De Martiis, Torino, ugo.colombino@unito.it.

2. The peculiar evolution of microsimulation

The first proposal for a «microsimulation model» appears in Orcutt (1957). More than an academic idea, it is a proposal addressed to the policy makers. At the time, the models used for policy analysis were macro models consisting of (mostly accounting) relationships among aggregates, with little and vague micro foundations. Orcutt's critique is focussed on four points: policy relevance, aggregation, micro foundations and representation of behaviour.

Policy Relevance

"*Existing models of our socio-economic system have proved to be of rather limited predictive usefulness. This is particularly true with respect to predictions about the effects of alternative governmental actions [...].*

Micro-foundations

[…] "*research efforts in the behavioral sciences have yielded and show promise of yielding very substantial amounts of knowledge about elemental decision-making units. However, existing models of socio-economic systems are neither built in terms of such units nor are they well adapted to making use of knowledge about such units*".

[…] "*The most distinctive feature of this new type of model is the key role played by actual decision making units of the real world such as the individual, the household and the firm*".

Aggregation

"[...] *current models of our socio-economic system only predict aggregates and fail to predict distributions of individuals, households, or firms* [...]".

"*Aggregation of relationships about elemental decision-making units is fairly easy if the relationships to be aggregated are linear [...]. However, if nonlinear relationships are present, then stable relationships at the micro level are quite consistent with the absence of stable relationships at the aggregate level*".

Representation of behaviour

"*This new type of model consists of various sorts of interacting units which receive inputs and generate outputs […] Probability distributions specify the probabilities associated with the possible outputs of the unit.*

"*Prediction about aggregates would still be needed but will be obtained by aggregating behavior of elemental units…*".

Summing up:

Policy-relevant models should be based on a disaggregated and explicit representation of the micro-units and of their interactions. Micro choices are represents as probabilistic events. The probabilistic representation of behaviour naturally suggests simulation as the tool to solve the model: thus, micro-simulation. A very ambitious project indeed: early improvements at the end of the 50s in micro data collection and management and in digital computing made Orcutt confident in the feasibility of the project. Realistically, the project had to be articulated in specific building blocks or modules.

Orcutt and associates proceeded to the implementation phase (Wisconsin, Urban Institute, Yale) working in particular on the household sector and on socio-demographic dynamics. A summary of early implementations is provided by Orcutt et al. (1961).

The story that follows can be divided into four periods: Conflictual marriage, Divorce, Preparing for a re-encounter, Re-marriage.

Conflictual marriage

In principle, Orcutt's proposal would have represented an ideal match between policy-relevant modelling and microeconomics (or micro-analytic behavioural theories in general). However, the two partners were not ready for that. Orcutt – a background in engineering and physics - had probably little confidence in microeconomic theory. Orcutt's project did not receive much interest either by microeconomists or by econometricians. Appropriate, policy relevant, empirical specifications of microeconomic models (i.e. microeconometric models) were not available yet.

As a consequence, the behavioural relationships illustrated for example in Orcutt et al. (1961) are reduced form specifications. This approach apparently contradicts Marshack (1953), whose lesson essentially tells that in order to be able to give policy prescription you need structural models or at least estimates of policy-invariant parameters. It must be noted that Orcutt wrote in a period were the empirical design of microeconomic policies was absent: therefore, the issues related to budget sets modifications implied by micro-policies were not on the agenda. In any case, probably Orcutt and associates thought that the most urgent destination of research efforts and resources was the exploitation of newly available micro data and computing resources.

Divorce

During the 70s, 80s and 90s, large microsimulation models in various countries (US, Canada, Scandinavian countries, Australia) acquire popularity, also at the policy making level.

The microsimulation community in this period focusses on the quality of data and the accounting reliability of the predictions. Behavioural responses are left outside. Non-behavioural models are more palatable to policy makers. Large part of the research effort is devoted to tax-benefit simulation models (e.g. EUROMOD).

Preparing for a re-encounter

During the same period that marks a divorce, many developments – at various levels: policy, theory, empirical methods – take place, preparing for a more mature and fruitful re-encounter.

- At the policy level, starting with the mid 60s, there is an increasing interest (war on poverty, tax reforms, welfare reforms etc.) in issues that involve structural changes in the opportunity sets.
- The lesson by Marshack (1953) and Hurvicz (1962) – revived by Lucas (1976) – i.e. you need structural models to make policy simulation, gets eventually fully learnt.
- Heckman (1974), Hausman (1985) and many others develop appropriate models to account for the complexities in the opportunity set (as those implied by newly conceived tax-benefit reforms).
- Discrete choice and random utility models (McFadden 1984) offer new and more flexible tools to estimate and simulate choices subject to complicated constraints.
- Applied microeconomists start using microsimulation techniques to compute responses to policies (Zabalza 1983: possibly the first one). Traditionally, even when using micro

data, economists used to compute behavioural responses for an “average” or “representative” individual and ignored the random components: a procedure that can lead to misleading results when the behavioural relationship is non-linear.

Re-marriage

The third millennium marks the re-encounter of microsimulation and microecono (-mics)(-metrics). The two partners are now ready.

- An example: Aaberge et al. (2000), presented at the workshop on Microsimulation in the New Millenium, Cambridge, 1999.
- In 2010, ISER (that hosts EUROMOD in Essex) organizes a large workshop on behavioral responses in microsimulation models.
- Compare the program of the IMA conference in Camberra 2003 (most of the papers are arithmetic) to the IMA conference 2013 again in Camberra (most of the papers are behavioural, especially labour supply) and to the last 2014 European IMA conference in Maastricht.

The current period witnesses a clearer vision of the respective roles of non-behavioural and behavioural models, of their possible integration and of new methods to extend the non-behavioural models. For decades, and for good reasons, policy makers remained suspicious about the reliability and generality of behavioural models. At the same time, they have learned to appreciate the value of non-behavioural microsimulation models as robust and invaluable tools. Now they start realizing that some representation of behavioural responses would be important and in some cases not dispensable. As an example, recent discussions about the redesign of income support mechanisms naturally lead us to ask about incentives (to work more or less or to work at all). Issues of this kind require in one way or another a representation of behavioural responses. We might say that static microsimulation models, besides their intrinsic value, have acted as a “benevolent” Troy horse in channeling the perspective of behavioural modelling.

However, this does not mean that the only way to take is a full integration of static and behavioural models. Especially from the point of view of an institutional research department, a cautious approach is certainly appropriate. First, under many circumstances, static microsimulation is all is needed. Second, even without adopting a full-blown structural behavioural model, there are various procedures to “enrich” the static microsimulation results and make it possible to produce approximate inferences on behavioural and welfare effects. In Section 4 we will provide a few examples.

3. Microsimulation models in Italy and in the World

I will start with a note on terminology. In the microsimulation community, there is some unnecessary ambiguity in the way different types of models are denominated. My preference would be to distinguish two dimensions: time (static vs. dynamic) and behaviour (behavioural vs. non-behavioural). This type of classification is consistent with the tradition established in economic theory. The analysis of how a consumer’s budget set changes due to changes in prices and or income is non-behavioural (although it might suggest some likely changes in behaviour as well: see Section 4). The same analysis, however, could be static (i.e. it might refer to a permanent scenario in a given period, whatever the length of period) or dynamic (i.e. it might refer to an intertemporal budget set). A static and non-behavioural analysis would investigate how the opportunities or the constraints change due, for example, to population's ageing or to some exogenous change in consumers' characteristics or environment. A behavioural analysis would instead include the change in behaviour as a response to the changed budget set (whether static or dynamic). Outside economics, the expression static behavioural analysis could probably sound weird since behavioural responses need time to materialize: however, what economists refer to in this case is the analysis of an equilibrium configuration of opportunity sets and choices in a given point or period of time. Comparative statics is therefore the analysis of different equilibria: they might take place in different point in time but the analysis is not dynamic, since it is silent upon what happens meanwhile (see Colombino 2013 for a static behavioural simulation procedure that is consistent with the concept of comparative statics). A behavioural dynamic analysis should tell us what happens at different points in time (not necessarily equilibria) that are in a real, not figurative, sequence. Summing up, the classification – with some example – would be as follows:

Table 1 – Microeconometric model classification

	Non-behavioural	Behavioural
Static	e.g. EUROMOD	e.g. ECONLAV
Dynamic	e.g. DYNASIM	e.g. CAPP_DYN

Within the category of behavioural models, we might want to further distinguish between structural models and reduced-form models. Structural models aim at representing choices as function of structural – i.e. policy invariant – parameters (Marschak 1953; Hurwicz 1962). For example, when representing consumers' choices, a structural model would permit a separate identification of preferences (by assumption policy invariant) and opportunity sets (which can be modified by policies). Reduced-form models (e.g. CAPP_DYN) represent choices as functions of parameters that are mixtures of parameters that in general do not allow to identify how they might be affected by policy changes (Lucas 1976). Reduced-form models can provide a very good approximation under the observed policy regime. They may also provide reasonable approximations under policy changes, but that very much depends on the characteristics of the policy changes, and the models' performance is difficult to judge ex-ante. Behavioural structural models are more often developed by academic researchers in view of the analysis of some specific issue – e.g. the static behavioural model of Aaberge and Colombino (2013) or the dynamic behavioural model of Todd and Wolpin (2006) – rather than as general-purpose platforms to be used within an institution. An exception is represented by ECONLAV, a static behavioural – and structural – model (De Luca et al. 2012).

In the microsimulation literature, alternative – and in my view confusing - terminologies are also used:

- Non-behavioural models are often called static or arithmetic.
- Models that are behavioural and/or dynamic, are often called dynamic;

A recent survey describing the design of static non-behavioural models is provided by Li et al. (2014). Previous surveys include Merz (1991), Sutherland (1995) and Citro & Hanushek (1991). For surveys focussed on Europe and on EUROMOD in particular, useful references are Sutherland (1995) and Sutherland & Figari (2013). Besides the new ISTAT models described in this volume, other static non-behavioural Italian models are surveyed by Curiel (2012).

Although we already have many static microsimulation models operating in Italy and a number of researchers have been using them also as algorithms matched to behavioural microeconometric models, the research effort illustrated in this volume is very welcome as a major step in establishing a sort of "official" platform adopting state-of-the-art methods and best-choice datasets.

4. Static models and policy analysis

Non-behavioural models – by definition – do not account for behavioural responses. However, there are cases when their outputs allows inferences on welfare effects. Moreover, they can be complemented with statistics that are sufficient to make local and approximate inferences on both behavioural and welfare changes, without adopting a full-blown structural behavioural microsimulation approach.

An analogy with meteorological simulations and prediction might be useful. While we have models that explicitly produce the probability of (say) rain with a full-blown structural approach, a more common (and possibly thought to be more robust) procedure, consists of using a model to simulate the basic physical data and then complement them with expert evaluations, previous estimates and relevant statistics in order to generate a prediction of the event 'rain'.

The basic case

The standard scenario where non-behavioural simulations may be sufficiently informative is when the policies or the reforms can be represented as marginal changes in prices $\mathbf{p}$ and/or in unearned income *y*. Let $\mathbf{p}'\mathbf{x} = y$ be the consumer budget constraint. Note that that the bundle $\mathbf{x}$ might include hours of work (with a corresponding negative price, e.g. - *w*). Let $V(\mathbf{p}, y)$ be the indirect utility function. Let us consider a marginal change $(d\mathbf{p}, dy)$. Then we have: $dV = [\partial V/\partial \mathbf{p}]'d\mathbf{p} + \mu dy,$ where $\mu \equiv \frac{\partial V}{\partial y}$ is the marginal utility of income. By applying Roy's Theorem (i.e. $[\partial V/\partial \mathbf{p}] = -\mathbf{x}\mu$) we get: $\frac{dV}{\mu} = -\mathbf{x}'d\mathbf{p} + dy.$ The right-hand side is the change in the budget, conditional on the pre-reform consumption bundle $\mathbf{x}$. The left-hand side is the monetary equivalent of the change in utility. Therefore, the result tells us that the change in the consumer's budget (i.e. the basic result produced by a non-behavioural simulation) is a money-metric measure of the change in utility.

Turning to the production side, let $\mathbf{q}$ be the prices faced by the (price-taker) firm and let $\pi(\mathbf{q})$ be the profit function. Then $d\pi = [\partial \pi/\partial \mathbf{q}]'d\mathbf{q} = \mathbf{x}'d\mathbf{q}$ (due to Hotelling's Lemma $[\partial \pi/\partial \mathbf{q}] = \mathbf{x}$). The total change in (money-metric) welfare would then be $dW = \mathbf{x}'(d\mathbf{q} - d\mathbf{p}) + dy$. The example clarifies the logic that can guide extensions of static models in view of policy applications. However, it might be of limited practical value. This is so, because of two reasons:

i) Taxes or subsidies applied under the current policy regimes might make it impossible to represent the budget constraint and the profit function in the same way as we did above.

ii) In general, policy reforms might involve non-marginal changes.

Even when facing these complications, there are a variety of methods by which we can enrich the static simulation results in order to make approximate inferences upon behavioural changes and welfare effects. We illustrate some of them below.

Harberger-type approximations

Harberger (1964) showed that, in a perfectly competitive market and under mild assumptions, the welfare effect $\frac{dW(t)}{dt}$ of a small change dt of a tax applied to good x can be approximated by $t\frac{dx}{dt}$. If the change in t is not marginal (e.g. a change from t_0 to t_1), we can integrate the above expression:

$$W(t_1) - W(t_0) = \int_{t_0}^{t_1} t\frac{dx(t)}{dt}dt$$

which in turn could be approximated as a sum of terms $t\frac{dx}{dt}$ evaluated at different values of t in the range (t_0 , t_1). We only need local measures of behavioural response. A textbook application is the "triangle" formula for the consumers' net welfare change:

$$W(t_1) - W(t_0) = -\frac{1}{2}b\left(t_1 - t_0\right)^2$$

where –b is the slope of a linear demand curve.

Chetty's "sufficient statistics" approach

Chetty (2009) generalizes Harberger's approach to more interesting cases (heterogeneous agents, discrete choices etc.). The idea essentially consists of complementing non-behavioural computations with "sufficient statistics" of local behavioural response, thus avoiding the need to develop a full-blown structural behavioural model. Note that these methods produce approximate results both on behavioural responses and on welfare effects. As a simple example, given a non-marginal variation of the wage rate, we could approximate the labour supply response by applying previously estimated labour supply elasticities (at the extensive and/or intensive margin, depending on the starting position of the individual). Saez (2001, 2002) has derived optimal tax-benefit formulas that only require local measures of intensive and extensive labour supply elasticities. (Immervoll et al, 2007) provide an empirical application. Of course there a price to pay when dispensing with the assumptions required by explicit behavioural simulation: the assumptions leading to the theoretical formulas and those underlying the empirical measures of elasticities are in general different and might not be mutually consistent (this problem carries over to Harberger's approach).

Local measures of incentives

Instead of computing local approximations of behavioural changes, one might simply compute local measures of incentives that are likely to induce changes (Jara and Tumino, 2013). Examples include the computation of Replacement Rates (Immervoll and O'Donoghue, 2004; O'Donoghue, C. 2011), Marginal Effective Tax Rates (Harding and Polette, 1995; Beer, 2003; Creedy et al., 2003; Scholz, 1996; Dolls et al, 2012) and the rate of return to education (O'Donoghue, 1999).

Using discrete opportunities as a menu of potential choices

A different line of attacking the problem consist of considering discrete opportunity sets. This comes natural for example when tax-benefit reforms and labour supply responses are the focus of interest. As an example, it might be natural to assume that each individual can choose among a (small) set of alternatives, such as non-working, part-time and full-time. For each alternative we can compute the net available income given a certain tax-benefit system. For individuals who are observed as not working, we will need to impute (with some missing-values-filling procedure) the gross wage rate. Then, for a generic individual, the alternatives could be described as follow:

$$\left(L_1, C_1^R\right),...,\left(L_M, C_M^R\right)$$

where L_j = "leisure" available if alternative *j* is chosen and C_j^R = net available income under tax-benefit regime *R* if alternative *j* is chosen.

Let *k* indicate the currently chosen alternative under the current tax-benefit regime *R* = 0. Standard presentations of results would for example consist in running the non-behavioural microsimulation model give the chosen alternative *k* under a new tax-benefit regime *R* = *P* and comparing, say, C_k^P to C_k^0. A more informative report would be produced by running the model at all the alternative 1, …, M. This would lead to comparing

$$\left(L_1, C_1^P\right),...,\left(L_M, C_M^P\right)$$

to

$$\left(L_1, C_1^0\right),...,\left(L_M, C_M^0\right).$$

Even without any explicit behavioural modelling or measures, the comparison might suggest likely directions of behavioural responses.

Evaluating discrete opportunity sets with "calibrated" utility functions

This further enrichment builds on the previous one and assumes that we are prepared to use some standard evaluation (utility) function $U = V(L, C) + \varepsilon$. This function could have been previously estimated. If we are not prepared to rely on demanding econometric estimates, the function *V* could be "calibrated". As a crude simple example, one might consider $V(L,C) = L^{\alpha} C^{(1-\alpha)}$ and "calibrate" the value of α for example as follows:

$$\alpha = L_k \big/ \left(T + y / w\right)$$

where
L_k = observed choice (hours of leisure)
T = total available time
y = gross unearned income
w = gross wage rate.

Note that one could compute a different α for each individual. If the random variable is assumed to have a standard extreme value distribution, then

$\ln \sum_j e^{V\left(L_j, C_j^R\right)}$ is a measure of average utility attained under regime *R*. Moreover,

$\sum_i C_i^R \dfrac{e^{V\left(L_i, C_i^R\right)}}{\sum_j e^{V\left(L_j, C_j^R\right)}}$ is a prediction (including behavioural responses) of the net available income under regime *R*.

5. Conclusion

From the viewpoint of database, methodology and scope for detailed policy analysis, the new Istat microsimulation models promise to be the most up-to-date official microsimulation platform. Besides the basic utilization as producers of timely non-behavioural simulation, they would also be very useful as algorithms matched to structural microeconometric models. Moreover, I suggested that occasionally they might also be complemented with "sufficient statistics" that, to a certain extent, permit inferences on behavioural and welfare effects without requiring an explicit structural behavioural model.

References

Aaberge, R. and Colombino, U., 2013. *Using a Microeconometric Model of Household Labour Supply to Design Optimal Income Taxes*. Scandinavian Journal of Economics. 115(2): 449–475.

Aaberge, R., U. Colombino, Strøm, S. and T Wennemo., 2000. *Joint Labour Supply of Married Couples: Efficiency and Distributional Effects of Tax Reforms* In: H. Sutherland, L. Mitton and M. Weeks (Eds.) Microsimulation in the New Millenium: Challanges and Innovations. Cambridge University Press.

Beer, G., 2003. *Work incentives under A New Tax System: the distribution of effective marginal tax rates in 2002*. Economic Record. 79 (Special Issue): S14-S25.

Chetty, R., 2009. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods". *Annual Review of Economics*, Annual Reviews. 1(1): 451-488.

Citro C.F. and E.A. Hanushek (eds.)., 1991. *The uses of Microsimulation Modelling*. Washington: National Academy Press.

Colombino, U., 2013. "A new equilibrium simulation procedure with discrete choice models". *International Journal of Microsimulation*. 6(3): 25-49.

Creedy, J., Kalb, G. and Kew, H., 2003. "Flattening the effective marginal tax rate structure in Australia: policy simulations using the Melbourne Institute Tax and Transfer Simulator". *Australian Economic Review*. 36(2): 156-172.

Curiel, A., 2012. *Modelli di microsimulazione delle politiche fiscali e previdenziali*. CaFoscari, Venezia, http://dspace.unive.it/bitstream/handle/10579/2533/986863-1167813.pdf?sequence=2.

De Luca, G., Rossetti, C. and Vuri, D., 2012. "In-Work Benefits for Married Couples: An Ex-Ante Evaluation of EITC and WTC Policies in Italy". *IZA Discussion Papers* 6739, Institute for the Study of Labor (IZA).

Dolls, M., Fuest, C. and Peichl, A., 2012. "Automatic stabilizers and economic crisis: US vs. Europe". *Journal of Public Economics*. 96(3): 279-294.

Harberger, C., 1964. "The measurement of waste". *American Economic Review*. 54(3): 58-76.

Harding, A. and Polette, J. 1995. "The Price of Means-Tested Transfers: Effective Marginal Tax Rates in Australia in 1994". *Australian Economic Review*. 28(3): 100-106.

Hausman, J. A., 1985. "The Econometrics of Nonlinear Budget Sets". *Econometrica*. 53(6): 1255-1282. doi: 10.2307/1913207.

Heckman, J. J. 1974., "Effects of Child-Care Programs on Women's Work Effort". *Journal of Political Economy*. 82(2): S136-S163.

Hurwicz, L., 1962. "On the Structural Form of Interdependent Systems". In: Nagel, E., Suppes, P. and A. Tarski (eds.) *Logic, Methodology and Philosophy of Science*, Stanford: Stanford University Press: 232-239.

Immervoll, H. and O'Donoghue, C., 2004. "What Difference does a Job Make? The Income Consequences of Joblessness in Europe". In Gallie D. (ed.) *Resisting Marginalisation:*

Immervoll, H., Kleven, H. J., Kreiner, C. T. and Saez, E., 2007. “Welfare reform in European countries: a microsimulation analysis”. *The Economic Journal.* 117(516): 1-44.

Jara, H. X. and Tumino, A. 2013. “Tax-benefit systems, income distribution and work incentives in the European Union”. *International Journal of Microsimulation*. 1(6): 27-62.

Li, J., O’Donoghue, C., Loughrey, J. and Harding, A., 2014. “Static Models”. In O’Donoghue C. (Ed.), *Handbook of Microsimulation Modelling*. Emerald.

Lucas, R. Jr., 1976. “Econometric policy evaluation: A critique”. Carnegie-Rochester Conference Series on Public Policy. 1(1): 19-46, Elsevier.

McFadden, D., 1984. “Econometric analysis of qualitative response models”. In Griliches†, Z. and Intriligator, M. D. (eds.), *Handbook of Econometrics*, edition 1, volume 2, chapter 24: 1395-1457, Elsevier.

Marschak, J., 1953. “Economic measurements for policy and prediction”. In Hood, W. and Koopmans, T. (eds.), *Studies in Econometric Method*: 1-26. New York: Wiley.

Merz, J., 1991. “Microsimulation- a survey of principles, developments and applications”. *International Journal of Forecasting*. Vol. 7, No. 1.

O'Donoghue, C., 1999. “Estimating the rate of return to education using microsimulation”. *Economic and Social Review*. 30(3): 249-266.

O'Donoghue, C., 2011. “Do Tax–Benefit Systems Cause High Replacement Rates? A Decompositional Analysis Using EUROMOD”. *Labour*. 25: 126-151. doi: 10.1111/j.1467-9914.2010.00501.

Orcutt, G. H., 1957. “A New Type of Socio-Economic System”. *Review of Economics and Statistics*. 39(2): 116-123. doi: 10.2307/1928528.

Orcutt, G., Greenberger, M.. Korbel, J. and Rivlin, A., 1961. *Microanalysis of socioeconomic systems: A simulation study.* New York: Harper & Row.

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”. *Quarterly Journal of Economics*. 117(3): 1039-1073.

Scholz, J. K., 1996. “In-work benefits in the United States: the earned income tax credit”. *The Economic Journal*: 156-169.

Sutherland, H., 1995. *Static microsimulation models in Europe: a survey*. Faculty of Economics, University of Cambridge.

Sutherland, H. and Figari, F., 2013. “EUROMOD: the European Union tax-benefit microsimulation model”. *International Journal of Microsimulation*. 6(1): 4-26.

Todd, P. E. and Wolpin, K. I., 2006. “Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility”. *American Economic Review*. 96(5): 1384-1417.

Zabalza, A., 1983. “The CES Utility Function, Non-Linear Budget Constraints and Labour Supply: Results on Female Participation and Hours”. *Economic Journal*. 93(37): 312-30.

Micromodelling Italian Taxes and Social Policies[1]

Maria Cozzolino[2], Marco Di Marco[3]

Abstract

Microsimulation models are nowadays extensively used to estimate the effects of existing and planned welfare and tax policies. The paper summarises the microsimulation models recently built at the Italian National Institute of Statistics (Istat) for the analysis of the Italian tax-benefit system and for the evaluation of public policies. It summarises the extent of Istat microsimulation models, focusing on the techniques applied to the available survey and administrative microdata to estimate the incidence of income, property and consumption taxes and of social security contributions, and the allocation of cash benefits. These calculations, taken as a whole, permit to assess the ultimate effect of the tax-benefit system on the redistribution of incomes and on poverty. A wide set of legislative details have been included in the models, in order to account for all the interactions between different policy instruments.

Keywords: Microsimulation, Redistribution, Poverty, Inequality.
C54 Quantitative Policy Modeling
H24 Personal Income and Other Nonbusiness Taxes and Subsidies
H53 Government Expenditures and Welfare Program

1. Foreword

As an independent public research agency, the Italian National Statistical Institute (Istat) is continuously asked to display its data and to provide advice to institutions, governmental bodies, researchers, universities and to general public on a host of different social and economic issues. For what concerns the monitoring of the Italian economy, the Parliament ordinarily consults Istat on the yearly Budget Law and on other regulations issued by the Government during dedicated sessions.

Moreover, the Institute has a special commitment to continuously support and advice the Research Department of the House of Deputies, providing *ex ante* and *ex post* evaluations especially on the effects of policies on the Italian economy, including the impact on the distribution of household incomes. Finally, Istat publishes regularly its research results in the Yearly Report and in other publications, press conferences etc..

---

[1] The opinions are those of the authors and do not entail the responsibility of their Institution. Although the article is the result of a joint work, paragraph 1 has been drafted by Maria Cozzolino and foreword and paragraphs 2 by Marco Di Marco and we are grateful to an anonymous referee who contributed with useful suggestions.

[2] Senior Researcher (Istat, Italian National Statistical Institute), e-mail: mcozzolino@istat.it
[3] Senior Researcher (Istat, Italian National Statistical Institute), e-mail: madimarc@istat.it

To meet such demanding goals, Istat has devised in the last two years an entirely new set of microsimulation tools including a model of direct taxes, social security contributions and cash benefits, a special module to simulate the real estate tax and a model of value added and excise taxes. The following list includes the social policies that are present in the surveys for the year *T* and/or can be made available through microsimulation techniques for any desired year *T + j* to provide *ex ante* and *ex post* assessments:

- Wage Integration Fund (*CIG*) allowances: ordinary and special
- Unemployment, mobility, early retirement indemnities
- Apprenticeship or employability allowances
- Scholarship allowances
- Lump sum related to unemployment
- Occupational pensions, that is, old age or seniority
- Disability allowances and pensions
- Pensions for occupational injuries or diseases
- Social pensions and welfare benefits for the low-income elderly
- War pensions (excluding those paid to survivors)
- Vouchers for disability to reimburse *in-kind* personal services (e.g . Taxi vouchers)
- Survivors pensions
- Supplementary pensions paid by private pension funds
- Value of the subsidized rent (when publicly owned and below the market price)
- Family allowances to employees and quasi-dependent self-employed
- Family allowances to retired
- Family allowances to laid-off workers and unemployed
- Maternity allowances, during the compulsory leave for employees
- Maternity allowances for women not entitled to the standard Maternity allowance
- Special allowances for families with at least three children
- Minimum Income provided by (some) local authorities to poor households
- Public contributions to housing costs (rent and/or interest payments on the mortgage)
- "Social Card" allowances for poor households (electricity and gas bills, food)
- Expenses deductible from the income tax (tax credits)
- Expenses deductible from the income tax base (tax allowances)
- Social security contributions on employers
- Social security contributions on employees and self-employed
- Personal Income Tax (Irpef), including local additional tax liabilities
- Tax expenditures (tax deductions and credits) related to Irpef
- Tax on Productive Activities (Irap) charged on the share of the labour income of the self-employed in the value added
- Municipal Real Estate Tax (IMU)
- Value Added Tax
- Excise taxes

Istat has built the new set of models on the basis of an established tradition in the use of microsimulation for policy evaluation, that has been enhanced by the merge with the Institute of Studies for Economic Analysis (Isae) in 2011. In the last three decades, both institutes had separately created their own micromodels. Isae pioneered microsimulation studies of the Italian tax-benefit system with Itaxmod, a static model built in 1988 and partly reshaped in the following years[4]. A dynamic model, Midas, has also been developed at Isae in 2009 in order to simulate future developments in the adequacy of pensions [5]. On its turn, since 1997 Istat maintained Mastrict, a static micromodel for the evaluation of policies relating to taxes, transfers, and social security contributions[6].

Both Mastrict and Itaxmod were based on the Bank of Italy Survey of Household Incomes and Wealth, whilst Midas was built on the microdata from the Europanel Survey. Also, both institutes have built micromodels of the Value Added Tax on the basis of the Istat Household Budget survey. More recently, Istat has contributed to the setup of SM2, a multi-country model devised by the University of Siena to perform the gross-net conversion of the incomes of the EU-SILC (European Union Statistics on Incomes and Living Conditions) project[7].

FaMiMod, the new Istat tax/benefit model is based on the EU-SILC dataset, which is now considered the best available source for the building of microsimulation models in the European countries. EU-SILC is jointly managed by Eurostat and the National Statistical Institutes of the EU countries and follows the definitions and best practices recommended by the Canberra expert group of the United Nation Department of Statistics (Canberra Group, 2001). Istat is in charge of providing the Italian version of EU-SILC.

Istat micromodels permit to evaluate *ex ante* the impact of tax and social policies on households, encompassing in the estimations new social programs and taxes as well as reforms of the existing ones. A crucial preliminary task of any micromodel is the projection of the information collected in year T to a later year $T + j$ for which no data are already available. In principle, this allows to simulate the effects of a change in policy at any given future date.

The average changes in the income variables are taken from the latest years of National Accounts data available or are foreseen by Istat macroeconometric model MeMo. The projection also requires a re-weighting procedure to account for changes in the demographic structure and in the employment status of the population (*static ageing*)[8]. The distributions of the updated/simulated variables are validated against the actual original data and current National Account figures (Figure 1). Visualization of the data permits a quick assessment of the main characteristics of the simulated variables.

4 Whilst Di Biase *et al.* (1995) describe Itaxmod, Sutherland (1995) reviews the initial spread of microsimulation studies in Europe, including Italy. The reader should be alerted that at the time Isae was named Ispe (Institute of Studies for Economic Programming).

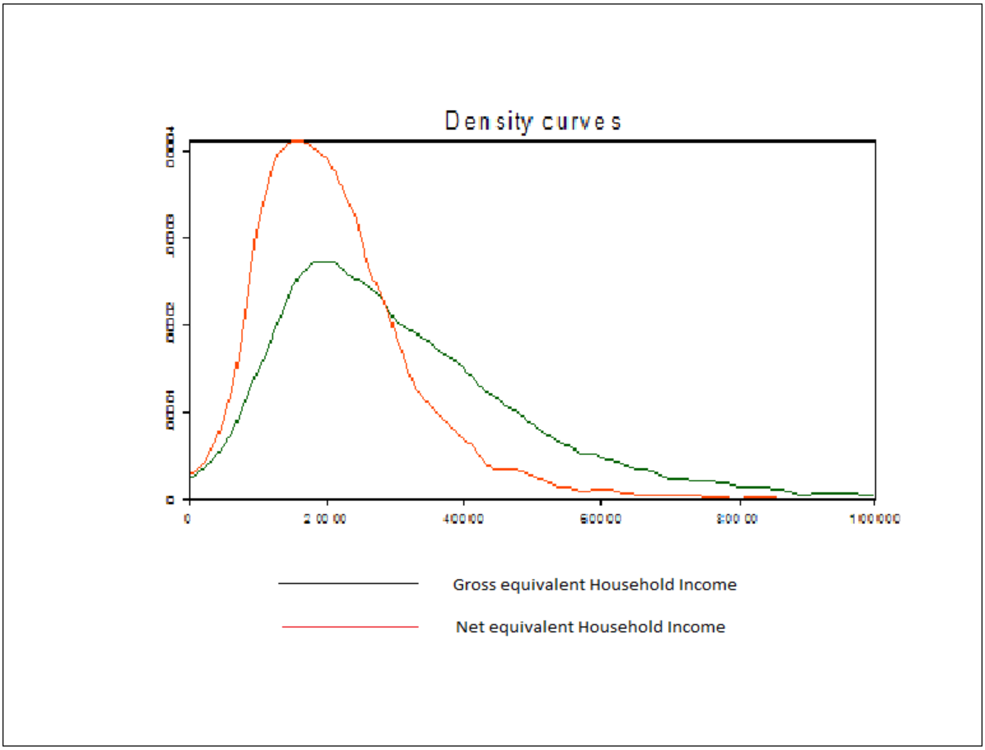
5 Dekkers *et al* (2010).

6 Proto (1999).

7 Betti, Donatiello and Verma (2011) and Istat (2011).

8 Alternatively, one could account for demographic changes with dynamic ageing, that "extend the static model by allowing individuals to change their characteristics due to endogenous factors" (O'Donoghue, 2001).

Figure 1 - Density curves of gross and net equivalent household income (simulated)

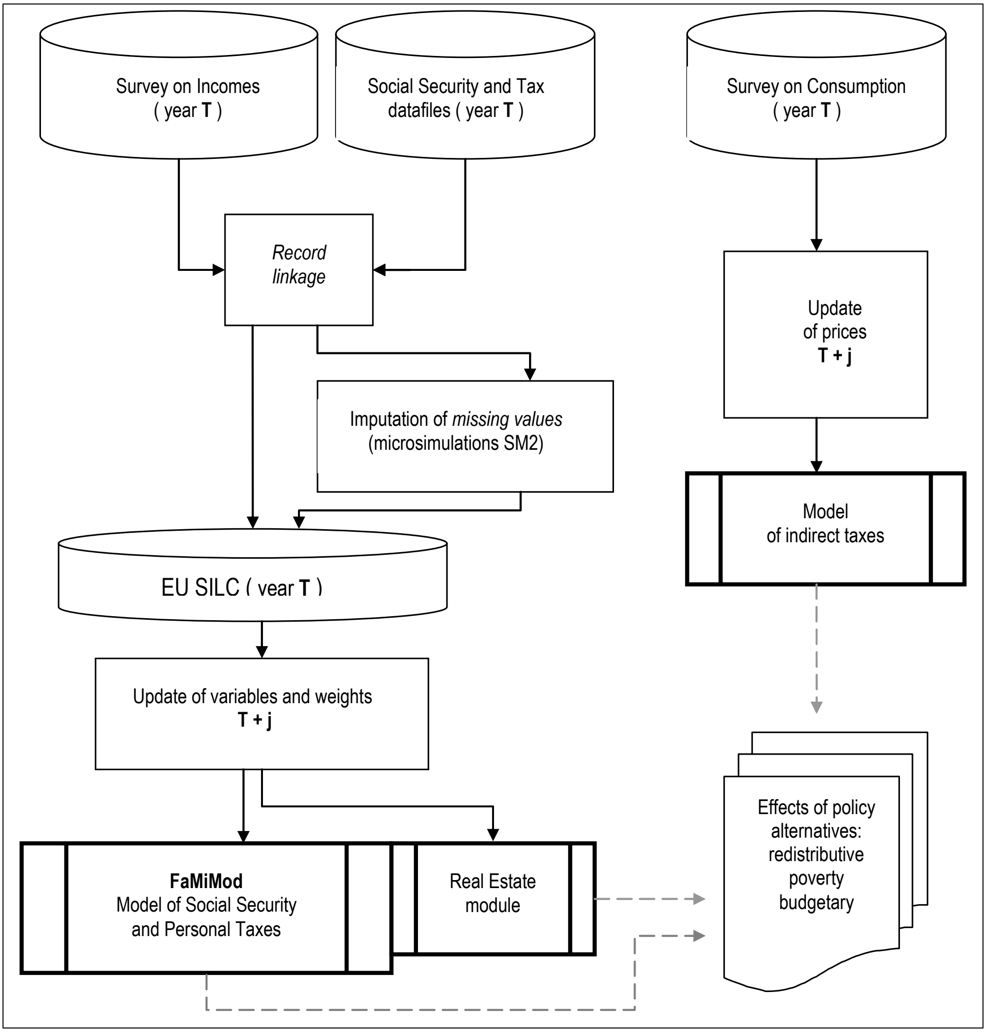


The Istat set of micromodels is based on two main sources of data: the EU-SILC (Income and Living Conditions) and the Household Budget surveys, both run by Istat (Figure 2). The EU-SILC survey is actually a blend of information collected directly from the households and of administrative data ‘imported’ from the tax and social security files through an exact matching (*record linkage*) at the individual level. Besides, the missing values of some variables are imputed through the Italian SM2, a microsimulation model specifically built for the EU SILC database. The characteristics of EU-SILC are described in depth in chapter 5 of this special issue of the *Rivista di Statistica Ufficiale*.

Chapter 2 of this special issue focuses on the bulding of FaMiMod, a microsimulation model of household taxes and benefits. The model accounts for taxes, social security contributions and cash benefits taken from the EU-SILC database of year T, adjusted and updated to the current (or future) year $T + j$. Besides, the model may optionally include in the gross (self-employment) income the share of the tax on “productive activities” (Irap) charged on the labour income of the self-employed in the value added.

The model has been implemented through three main phases: i) the construction of the database adjusted to year $T + j$, including the update and reweighting; ii) the development of a set of computer programs to simulate cash benefits, taxes and social security contributions in two alternative scenarios to be compared: the first reflecting the actual legislation and the second the modified rules; iii) the effects on the budget, on the distribution of incomes and on poverty are, simply, the differences in the available households incomes in the two alternative scenarios. Finally, to display the results the models include routines to compute statistics, indicators and graphics.

Figure 2 – Flow-chart of the Istat microsimulation models



Chapter 3 illustrates the special module setup to estimate the Municipal Real Estate Tax (IMU) for the main dwelling owned by sample households, using the rates and the tax abatements actually applied in 2,599 Italian Municipalities. When a taxpayer who lives in his/her own house has no revenues subject to the income tax, or only perceives an employment income and/or a pension and has no need to file a tax return, the tax file does not report the taxable income from home-ownership. These missing values have been imputed through a regression on the responding home-owners who have to report the rental value of their main dwelling in the tax files.

Finally, chapter 4 describes the model on VAT and excise taxes used to measure the effects of changes in indirect taxes, assuming a complete translation onto consumer prices. The results provide: i) estimates of the impact of changes in the VAT rates on household expenditure, broken down by household types; ii) aggregated budgetary effects. A planned development is to match statistically the EU-SILC with the Household Budget surveys, to obtain an integrated income-expenditure set of microdata that will permit to assess the impact of indirect taxes with respect to household incomes.

2. Microsimulation models and the evaluation of public policies

Microsimulation models have come to be considered the most appropriate tools for the assessment of the redistributive effects of public policies, especially when these consists in the provision of cash transfers and other benefits or in the taxation of incomes, wealth or expenditures. Microsimulation is clearly more informative than the ‘typical agent’ approach, that is based on a few selected examples. Using information about thousands of individuals and households, microsimulation models permit to better identify who are the winners and losers of a policy and to estimate their number. Therefore, the overall effects (on equality, on poverty, on the budget etc.) can be assessed on the basis of a representative sample of the population. The policymaker concerns in the field of social benefits and tax policies should not be limited to a mere list of welfare indicators and accounting costs relating to each single policy instrument. Rather, the debate focuses, or should focus, on the efficiency and effectiveness of a single policy and/or of the tax/benefit system as a whole in preventing poverty and social exclusion. At this regard, one may note that even if a single policy is under evaluation, it is appropriate to account for all the interactions it has with all the other programs.

In the literature, efficiency and effectiveness issues are in fact addressed by analysing different questions:

Flaws in the design of policies

- **Leakage** is a typical symptom of the inefficient selection of the beneficiaries. It occurs when a part of the total expense for the benefit goes to people that the norm does not envisage as the ‘needy’, because of frauds as well as of mistakes in the design of the policy and/or in its actual administration (namely, of the access criteria). Leakage can be measured by the share of total expenditure that *does not* reach the "desired" target population.
- **Coverage**: how many of the ‘needy’ are actually entitled to a specific benefit? And, in case of under-coverage, how much money is needed to reach the whole target population? The so-called *take-up* problem (i.e. the self-selection of potential beneficiaries) explains part of the under-coverage and is usually due to lack of information and/or assistance to the applicants. However, under-coverage may arise because of an inadequate design of the policy, too. A pos-

sible measure is the percentage of percipients of a particular cash benefit on the total target population.

- **Sufficiency** of the measure with respect to the intensity of the percipients needs. For example, in the case of a cash subsidy for the poor, the extent to which it reduces the poverty gap (i.e. the difference between household income and the poverty line). The interesting measures here are the percentage of the gap covered by the policy and the amount of money needed to fill the gap completely.

All these aspects can be addressed to explore the effects of a single policy and of the whole tax-benefit system as well. However, the related measures can be estimated only if the researcher has sufficient and reliable administrative and/or survey microdata and a (set of) detailed microsimulation model(s).

Consistency between the general scopes and the effects of policies

A more general question is whether the set of all the taxes and benefits is consistent with the general scope of the system. The most explored issue, in this regard, is whether the system performs well in fighting poverty and social exclusion, usually measured with reference to an income threshold (i.e. a poverty line) or by other suitable welfare and well-being indicators. There is abundant literature on the pros and cons of the different definitions of poverty (relative and absolute) and of alternative measures of well-being.

The set of indicators chosen by the EU for the program Europe 2020 includes the (risk of) *relative* poverty, a measure of severe material deprivation due to the lack of some basic necessities and a low employment level of the household labour force. It is interesting to note that all these indicators are measured at the micro level, so that the *ex-ante* and *ex-post* assessment of the national policies is bound to rely upon measures of the effects at the micro level with collected data and/or micromodelling.

Usually, only few social policies meet all the requirements entailed by the many possible indicators of welfare, well-being and poverty. Moreover, it is very difficult to define commonly accepted measures of equality: each available welfare index, according to its mathematical properties, implies a different degree of aversion to inequality and weights differently the gains and losses of the rich and the poor. In principle, micromodels can provide a full set of indicators and thus indicate whether a policy change improves welfare according to each of the different measures available.

Another complex technical issue is the selection of a suitable benchmark for the comparison, that is the counterfactual scenario describing what might have happened had the policy under evaluation *not* been in force. The *coeteris paribus* clause is standard practice, provided that the possible interactions with other measures are taken into account in the simulations. Also, it is sometimes desirable to insert appropriate behavioural parameters in order to assess the individual reactions to the policy.

Comprehensiveness of the assessment of the effects

In principle, the evaluation of a social policy should not disregard externalities and non-monetary benefits. For example, child allowances for poor households may have positive effects on the human capital of the future cohorts of the labour force through higher education expenses that could bring about higher future increases in productivity of the economic system.

Similarly, it would be of paramount importance to understand whether countercyclical social policies concur to stabilize the macroeconomic fluctuations of the output. The standard practice in microsimulation exercises is to ignore all the potential macroeconomic effects of social policies and of tax policies, except for the *coeteris paribus* impact on the public budget. This limitation could only be avoided by integrating macro and micro models.

Finally, in light of the debate that goes under the title "beyond GDP", it is necessary to discuss on a case by case basis whether the evaluation of a policy should be conducted solely in terms of its *monetary* effects, including the exit from monetary poverty, or it is more appropriate to refer to the quality of life in its broadest sense (e.g. to non-monetary indicators of health, empowerment, quality of social capital etc). Again, this calls for further improvements of the microsimulation techniques. Namely, it would be necessary to encompass the correlations between the income levels, the extent of social policies and the observed non-monetary indicators of well-being at the individual level.

Estimating welfare effects with "arithmetical" microsimulation models

Most microsimulation models ignore the behavioural responses of the individuals to the change in the tax and/or benefit policies. These "arithmetical" models substantially compute the change in the disposable incomes at the individual or household level by comparing their budget before and after the new policy. This is equivalent to ignore any possible behavioural reaction (of the labour supply, of the household composition etc.) induced by the policy.

The assumption of constant behaviour is not too naive when it is important to measure the impact of a policy change in terms of its welfare effect. Applying the theory of consumer behaviour, the effect of a change in the bud**g**et constraint can be assessed on the basis of a money metric utility measure that evaluates the change in welfare as an "equivalent" variation of income. The tax-benefit policies change the price of the goods that a household consumes, of the services it sells on the market and/or its exogenous income. The assessments provided by "arithmetical" microsimulation models corresponds to the measurement of the equivalent variation that is obtained by applying the new prices and the new budget constraint to the initial consumption bundle and labour supply of the household[9]. This argument, however, does not overcome the problem of incentive compatibility. Effectively, in designing public policies aimed at changing individual behaviours to improve welfare (e.g., higher employment) , the incentives of individuals should be consistent with those of public and private agencies managing the public programs. Besides, when the expected effects of planned reforms are hardly attributable to marginal variations, microsimulation analyses need to go a step beyond a mere arithmetical perspective.A microsimulation model permits to identify the individuals and households who gain or lose money because of a policy and the monetary effect as well. A typical display of the results is the average gain or loss broken down by deciles of income, by socio-demographic characteristics, geographical area,

[9] For an algebraic proof of this equivalence see Bourguignon and Spadaro (2006)

specific target groups etc. A host of social welfare indicators can be computed to assess the inequality of the distribution of incomes before and after (with or without) the policy as well as the impact on poverty. Finally, one may calculate measures of progressivity and of polarisation/discrimination.

References

Betti, G., G. Donatiello and V. Verma (2011), The Siena Microsimulation Model (SM2) for net-gross conversion of Eu-Silc income variables, *The International Journal of Microsimulation,* 4(1): 35-53.

Bourguignon, F. and A. Spadaro (2006), Microsimulation as a Tool for Evaluating Redistribution Policies*, Journal of Economic Inequality*, 4, 1, 77-106.

Canberra Group (2001), *Expert Group on Household Income Statistics: Final Report and Recommendations*, Ottawa.

Consolini, P. (2009), *Integrazione dei dati campionari Eu-Silc con dati di fonte amministrativa*, Metodi e Norme n. 39, Istat, Roma.

Creedy, J. (2003), Survey Reweighting for Tax Microsimulation Modelling, *Treasury Working Paper Series* 03/17, New Zealand Treasury.

Dekkers, G., H. Buslei, M. Cozzolino, R. Desmet, J. Geyer, D. Hofmann, M. Raitano, V. Steiner, S. Tedeschi, F. Verschueren (2010), What are the consequences of the European AWG-projections on the adequacy of pensions? An application of the dynamic micro simulation model MIDAS for Belgium, Germany and Italy, in O'Donoghue C. (ed.), *Life-Cycle Microsimulation Modelling: Constructing and Using Dynamic Microsimulation Models* LAMBERT Academic Publishing.

Di Biase, R., M. Di Marco, F. Di Nicola and G. Proto (1995), ITAXMOD: A Microsimulation Model of the Italian Personal Income Tax and of Social Security Contributions, *Documenti di Lavoro ISPE*, n.16/95, Ispe, Roma.

Istat (2011), *La metodologia di stima dei redditi lordi nell'indagine EU-SILC*, Metodi e Norme n. 49, Istat, Roma.

O'Donoghue,C., Li,J. (2013), A survey of dynamic microsimulation models: uses, model structure and methodology, in: *International Journal of Microsimulation,* 6(2), p. 3-55.

Proto, G. (1999), Il modello di microsimulazione MASTRICT: struttura e risultati, *Rivista di Statistica Ufficiale,* n. 3, Istat, Roma.

Sutherland, H. (1995), Static Microsimulation Models in Europe: A Survey, *Microsimulation Unit Discussion Paper*, MU9503, Cambridge (UK).

Modelling Social Security, Direct Taxes and Cash Benefits [1]

Maria Cozzolino[2], Carlo Declich[3], Marco Di Marco[4], Gaetano Proto[5], Fabrizio Solari[6], Paola Tanda[7]

Abstract

The FaMiMod microsimulation model uses data gathered under the EU-SILC (European Union Statistics on Income and living conditions) framework, including both survey and administrative data, to assess the distributive effects of existing or hypothetical fiscal policies by computing taxes, social security contributions and some cash benefits at the micro level. The model estimates the main parameters of the distribution of individual and household incomes as well as aggregate welfare, equity and poverty indexes. Information on income referred to the year covered by EU-SILC is updated to current year by two procedures: (i) applying the National Accounts average growth rates of personal incomes, broken down whenever possible by source, geographical area and economic sector of activity; (ii) adapting the weights to account for changes in the demographic structure and in the employment status of the resident population occurred after the survey year. Social security contributions paid by both employees and employers are modelled, leading to an estimation of the current "tax wedge" on labour cost, as are the personal income tax, with special attention to the details of tax credits, and local taxes (regional and municipal) which have increased their weight in recent years. The main programme of family allowances is also modelled, in order to simulate redistributive policies that act simultaneously on the tax and on the benefit side.

Keywords: microsimulation, calibration, tax wedge, personal income tax, family allowances. JEL: C54 Quantitative Policy Modeling; D31Personal Income, Wealth, and Their Distributions; H71State and Local Taxation, Subsidies, and Revenue; I38 Government Policy Provision and Effects of Welfare Programs

---

1 The opinions expressed in this paper are those of the authors and do not entail the responsibility of their Institution. Although the article is the result of the joint work of the group of researchers, paragraphs 2 and 5.2 have been drafted by Gaetano Proto; paragraph 3 has been prepared by Fabrizio Solari; paragraph 4.1 by Maria Cozzolino and Paola Tanda, paragraphs 4.2 and 5.3 by Paola Tanda; paragraph 5.1 by Marco Di Marco; paragraph 6 by Carlo Declich. The authors are grateful to Corrado Pollastri and Floriana D'Elia for their valuable cooperation and to an anonymous referee who contributed with useful suggestions.

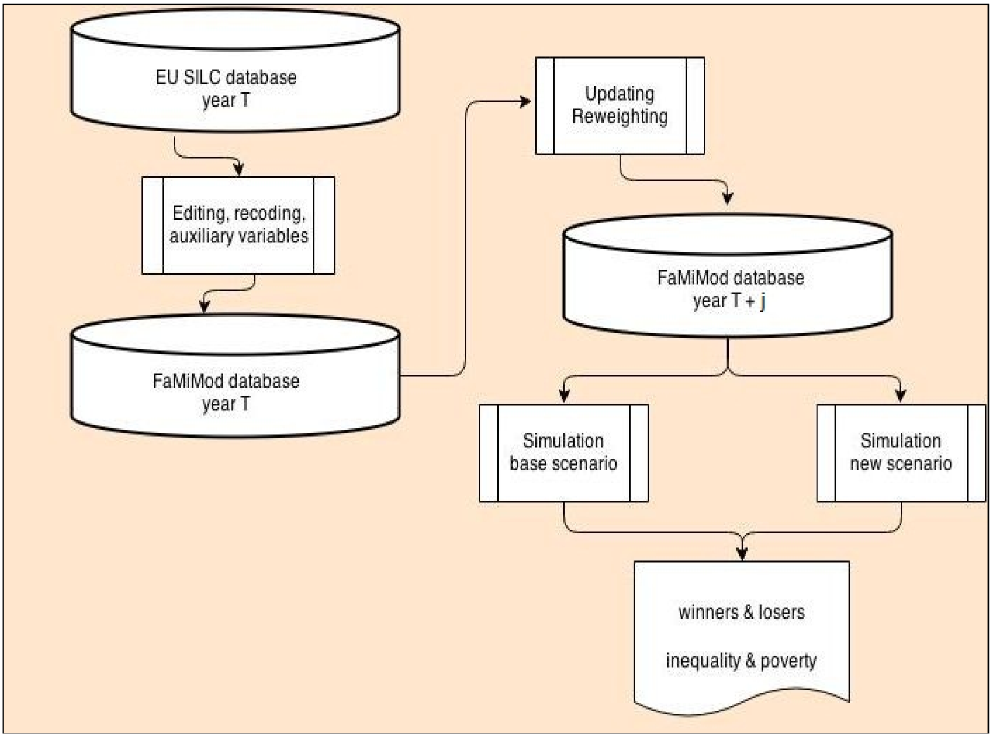
2 Senior Researcher, Istat (Italian National Statistical Institute), e-mail: mcozzolino@istat.it
3 Researcher, Istat (Italian National Statistical Institute), e-mail: declich@istat.it
4 Senior Researcher, Istat (Italian National Statistical Institute), e-mail: madimarc@istat.it
5 Senior Researcher, Istat (Italian National Statistical Institute), e-mail: proto@istat.it
6 Researcher, Istat (Italian National Statistical Institute), e-mail: solari@istat.it
7 Senior Researcher, Istat (Italian National Statistical Institute), e-mail: ptanda@istat.it

1. Introduction

FaMiMod, the microsimulation model of household taxes and benefits developed at Istat is a non-behavioral tax-benefit microsimulation model, as it does not account for the behavioral reactions of the individuals to the changes in the tax-benefit policies. Therefore, it is suited to assess the impact of the tax-benefit system and its changes on the distribution of personal and household incomes, focusing on short-term policy measures characterized by variations in policy parameters that are not too large. *Ex ante*, in the definition phase of the policy, the microsimulation model is a useful tool to assess the impact of the intended measures and to identify potential beneficiaries. The model is also useful for *ex post* evaluations of the effects of policies that have been undertaken, in order to assess the opportunity of corrective measures.

The structure of FaMiMod is based on a set of subsequent and interconnected modules, summed up in Figure 1. The first module, containing the selection of relevant variables from the data sources to be included in the database for the reference year and the creation of a set of new variables useful for the functioning of the model, is summarized in the following section, while the module describing the process of updating and reweighting to the current year is presented in section 3.

Figure 1 - Flow chart of the FaMiMod model



After these steps that lead to the updated database of the model, the modeling of the tax-benefit system takes control. The modules dealing with social security contributions and the subsequent calculation of the tax wedge are described in section 4. The personal income tax modules follow in section 5, including specific modules for the calculation of tax credits and of regional and municipal additional income taxes. Finally, the module describing family allowances is presented in section 6 while section 7 draws some conclusions and further developments.

2. Building the dataset for the FaMiMod microsimulation model

The database for the first release of the FaMiMod model builds on the 2010 edition of the Istat survey “European Statistics on Income and Living Conditions” (EU-SILC, income reference year 2009) [8]. This is a major innovation with respect to the previous microsimulation models based on the Bank of Italy survey on Income and Wealth (SHIW) such as MASTRICT and ITAXMOD (Di Biase et al. 1995, Proto 1999). In this section, the main steps of the building of the dataset will be recalled, pointing out when the new database has allowed to improve or at least to maintain the state of the art achieved by previous microsimulation models.

First of all, some basic preliminary choices had to be made regarding the unit of analysis, the definition of income and its disaggregation. The appropriate use of the available administrative information was also addressed in the planning phase.

Primarily, the main unit of analysis of FaMiMod is the individual. This is consistent with the individualistic nature of the Italian personal income tax, the most important component of the microsimulation model. This choice entails the reconstruction of some variables, as the survey total income is exhaustive only at the household level. Compared to the SHIW, where some income components are attributed to the head of the household, EU-SILC requires supplementary attention, because at the individual level misalignments may arise for some income components between net and gross variables that need to be detected and corrected. For capital incomes, the tax due on some components of income may have been attributed to a different member of the household than the one who actually perceives the corresponding net income, usually the one with the highest income, as the corresponding gross target variable is requested only at the household level. In the case of incomes from financial assets, the reallocation of taxes available at the household level in proportion to the reported net financial income of each member overrides these potential inconsistencies, and it will be a future refinement of the model to rely on more complex hypotheses, in order to account for individual portfolio structures and differences in the statutory rates applied to particular financial incomes.

Whilst the choice of the individual as the unit of analysis already was the standard of previous microsimulation models based on the Bank of Italy SHIW, in FaMiMod the choice regarding reference income has been an innovation. In fact, the SHIW only supplies information about the incomes net of taxes and social security contributions, while since the

[8] Other recent Italian tax-benefit microsimulation models based on EU-SILC are: ITaxSIM at the Department of Treasury, MAPP at the University of Modena and Reggio Emilia-CAPP (Baldini et al., 2011), MICROREG at the Tuscany Region-IRPET (Maitino et al., 2013). The Italian section of EUROMOD, a cross-national microsimulation model that produces comparable results for 27 EU countries, is also based on the same data source (Sutherland and Figari, 2013).

2007 edition the Italian EU-SILC includes gross incomes and therefore permits to take the *before tax and social contributions* income as the starting point of the updating and estimation procedures.

Besides, the Italian version of the EU-SILC survey is based on the integration of the survey data with administrative archives, the general principle being the building of gross incomes target variables on the basis of the net values reported to the interviewers, which may in turn be replaced by administrative data when these appear to be more complete, plus the taxes recorded in the tax files. Therefore, the logical structure of the new microsimulation model is simpler and more efficient with respect to pre-existing models, as the initial stage where net incomes are "grossed up" is no more necessary[9].

Another choice that is common to FaMiMod and to other models concerns the detailed level of disaggregation of income. In the context of a microsimulation model, indeed, total income cannot be treated as a single summary variable but must be processed as a collection of different components, each of them carefully checked for consistency and completeness.

A preliminary check has to be made on the consistency between survey variables and their definition in the tax code. First of all, net value and gross value must be identical for tax exempt incomes, such as the allowance for the attendance of disabled persons ("*Indennità di accompagnamento*"). In the 2010 edition of EU-SILC, taken as the basic source for the first release of FaMiMod, inconsistencies in this respect have a low incidence and the few observed cases could easily be corrected aligning gross to net values. On the other hand, the check of survey data against the maximum statutory values that had to be performed on the SHIW for some public transfers, like the welfare pensions to the elderly (*"Assegno sociale"* and *"Pensione sociale"*) and the consequent re-classification of the self-reported amounts exceeding the maximum statutory values are no longer necessary, thanks to the integration of the EU-SILC survey with the social security files.

As regards completeness, some variables missing or incomplete in the survey may turn out to be instrumental in the simulation of relevant details of the Italian tax law. For instance, information about the months worked by contract workers (termed Co.co.co or Co.co.pro, that is *Collaboratori coordinati e continuativi* or *a progetto*) is necessary to estimate the employment tax credit they are entitled to, which is approximately proportional to the number of days worked, since they are treated like employees for personal income tax purposes. Yet the months worked are not directly asked in the EU-SILC questionnaire (unlike the SHIW) as contract workers appear in the self-employed section, and they cannot always be retrieved from the answer to the question about month-by-month main activity in the income period, especially when they correspond to secondary activities. Thus, for the self-employed the missing values of the number of months worked have been estimated on the basis of gross income, assuming an average monthly reference income.

Disaggregation of income also permits the accurate modelling of differential tax regimes ('separate' taxation), of the exemption of specific income components as well as specific tax rules, such as the mentioned treatment of self-employment incomes of contract workers alike employment income. Along with the choice of gross incomes as initial inputs

[9] For a comparison of EU-SILC gross income data with the corresponding variables, simulated by subjecting EU-SILC net data to the net-to-gross module of a microsimulation model, see Ceriani, Fiorio and Gigliarano (2013).

of the model, the detailed disaggregation of incomes requires a supplementary effort in the proper treatment of the many variables used in the microsimulations, as the EU-SILC survey supplies more information on the components of net, rather than gross, incomes.

To this end, different procedures were followed. The simplest is to apply to the aggregate gross variable the ratio of the component, computed with reference to the aggregate net variable, in order to derive missing components of individual gross income starting from existing components of net, as in the cases of income from contract work with respect to total gross self-employment income and of arrears with respect to total gross labour income of employees. For severance pay, which is subject to separate taxation, the creation of a gross variable matching the net one asked for in the interview involves a plurality of income components, since severance pay can be allocated to unemployment benefits or to pensions (both old age and survivors) target variables in EU-SILC, depending on which social protection function it can be attributed to. Eventually, gross severance pay is the sum of three addends separately derived applying the specific ratio-of-net to the respective aggregate gross variable[10].

More complex hypotheses may be involved in the derivation of missing individual components of gross income from household components of gross income, as in the case of rental income: if co-owners are present, taxes paid on rental income available at the household level have been allocated to individuals as a proportion of their estimated average tax rates, since rents are included in taxable income.

Finally, a further innovative choice made possible by the use of EU-SILC as basic source is the exploitation of individual (anonymous) information stemming from tax records as a supplement to sample data. Whenever possible, tax records are linked to sample units by EU-SILC in the process of derivation of gross income target variables: some pieces of information have been made available for use to the team building FaMiMod, in order to fill in the gaps that sample surveys on incomes usually lack in the overall tax position of taxpayers.

This applies to expenses entitled to tax credits. In pre-existing models, health and other deductible expenses were imputed at their mean cell values based on occupation and income class, derived from semi-aggregated tax record data of previous years. This entailed a loss of individual variability that actual tax micro-data incorporated in the database of the FaMiMod model have allowed to avoid.

EU-SILC provides little information on real estate apart from owner-occupied housing, unlike the Bank of Italy SHIW which has wealth among its main objects of inquiry. Land and buildings other than home are lumped together in a single question, as are any rental incomes derived from them. Comparing sample data with information on income from land and buildings stemming from tax records, the latter showed a much broader coverage. Administrative data have therefore been used as a substitute, leaving to the future the possibility to complement them with sample information that would represent undeclared incomes.

Still regarding income from real estate, a further value added of tax records is to incorporate information concerning cadastral values, which otherwise could only be

[10] In fact, both arrears and severance pay are usually subject to separate taxation, providing for the application of an average estimated tax rate rather than the marginal rate: in this case, our simplified procedure of reconstruction entails some overestimation of their gross value.

approximately estimated on the basis of the respondents' self-assessed imputed rents of owner-occupied dwellings, as was usual with pre-existing microsimulation models. Nevertheless, information stemming from tax records may not always be exhaustive on its turn: this is the case of cadastral rent of owner-occupied housing, a part of the so-called "overall income" ("*reddito complessivo*") for income tax purposes until 2011. The value of this rent cannot be obtained from tax records if the individual is not obliged to file a tax return, either because he has no other income from real estate exceeding a low no-tax threshold, or because he has nothing to add to the record filled by his withholding agent. In these cases, it is still necessary to resort to imputations.

3. Reweighting and updating to current year

Typically, microsimulation models use a set of individual and household micro-data collected over a period *t* to estimate distributional, budget (and possibly behavioural) effects of changes in the taxes and benefits system in a subsequent period $t + j$, comparing them with the effects of the existing system. This type of analysis requires techniques allowing to incorporate into microsimulations at least some of the characteristics of the population of period $t + j$, which may be different from those of the reference population of the available micro-data, normally referred to a previous year.

The details of the techniques used by different microsimulation models to update the database to the simulation year $t + j$ depend on the particular purpose of each model and, specifically, on the characteristics of the data base and the availability of additional information about the external characteristics of the population at time $t + j$. It is therefore not easy to provide a general theoretical framework of methodological aspects related to temporal data updating for microsimulation models, although some contributions for specific models illustrate systematic problems related to the different techniques used (see, e.g., Creedy, 2003, or Immervoll et al., 2005).

A first general distinction can be established between:

- static updating techniques, which basically consist in the recalculation of the individual and/or household weights, in order to replicate the known totals of some characteristics of the population at time $t + j$;
- dynamic updating techniques, typically based on micro-econometric models, which estimate the probability of transitions, i.e. changes in the individual life cycle (e.g., births, deaths, marriages/divorces, inputs/outputs from the labour market, etc.).

It is also useful to distinguish, conceptually, with respect to these two types of methods related to representation at time $t + j$, the values of the analysis variables (e.g. income). In fact, they are two separate issues: the demographic representativeness of a random sample of individuals after a certain period of time depends on the inputs/outputs from the labour market by layoffs and hiring, from births and deaths, as well as the processes of internal reorganization (for example, if managers have increased with respect to the employees, women with respect to men, changes in the economic activity sector, etc.).

These changes may work independently of the changes in the characteristics of individual occupations, i.e. wages, hours worked, etc., and it is therefore common practice in the construction of microsimulation models to update the database both for the individual characteristics and for changes in income (or other variables such as hours worked, taxes, etc.). In both cases, the distribution of analysis variables (such as income) is altered with respect to the base year. Ideally, to replicate the income distribution at year $t + j$ it would be desirable to have the highest level of detail on the variation of analysis variables, for example at the level of individual statistical unit, and the reference population distinguished by characteristics such as gender, age, employment status (particularly important are those related to changes in income). In practice, for the variation of analysis variables the most recent aggregate data (usually those from the National Accounts) actually allow to assign only mean changes to particular sub-groups (for instance the sub-groups of earners defined by different income sources) and therefore represent only one of the possible changes of the distribution occurred between time t and time $t + j$.

The year initially chosen for the representation of the system of taxes and benefits is 2012. The update from the base year is a basic feature of microsimulation models that allows to project the available information to a year for which final data from surveys or National Accounts are usually not yet available, especially when the model is used to evaluate current or planned policy measures. The update from 2009, the period for which we have information on incomes, to 2012 has been realized by applying the changes in the variables of interest using the various available sources, primarily the National Accounts. When aggregate data was not available, the macroeconomic forecasts provided by Istat macroeconomic models have been used. The projection was made considering the possible articulations by sector of activity and sources of income. Most pensions have been updated taking into account the normative concerning the adjustment to actual inflation.

Assuming that these average changes are sufficiently informative of the situation at time $t + j$ of individuals “represented” by the sample observed at the time t, the recalculation of the weights tries to make the sample units at time t consistent with the characteristics of the population available at time $t + j$. Calibration techniques (Deville and Särndal, 1992) may be used to adjust the sampling weights available for the sample data at time t. This is done by minimizing the differences between the weights at time t and the weights to be used for time $t + j$, and under the condition that the final weights computed for time $t + j$ must satisfy a set of constraints related to known characteristics of the population at time $t + j$.

Note that the new weights can in principle capture a part of those changes of the income distribution between time t and $t + j$ that cannot be represented through the average increase of analysis variables, replicating the composition effects. The most important of these effects are related to demographic changes, such as the distribution of the population by gender, age and employment status. The population counts of these variables are then considered for the adjustment of the weights available for time t. More controversial are the effects on the simulated distribution of any constraint corresponding to total amounts of some variables of investigation, such as income, which depends on a set of factors that are not perfectly kept under control by researchers, as for example the bias due to partial or total non-response.

As stated above, the data source used in the FaMiMod microsimulation model is the Italian EU-SILC survey. The survey is carried out yearly after the deadline to file tax returns in order to give to households and individuals the opportunity to use the information resulting from them. The survey provides both cross-sectional data on income, poverty and social exclusion and living conditions, and longitudinal data on income, employment and some non-monetary indicators of social exclusion.

The survey is based on four longitudinal samples. These samples are shifted in time so that in each wave take place the closure of the panel arriving to the fourth wave and the beginning of a new panel. Each longitudinal sample is the output of a two-stage stratified sampling design, with stratification of the primary sampling units, the municipalities, while the secondary sampling units are the households. The cross-sectional sample is composed by the union of four longitudinal samples, each for its specific wave.

The whole weighting procedure for the EU-SILC sampling design is extremely complex, since it covers both cross-sectional and longitudinal development. Therefore, the procedure can be divided in different sections according to their purpose.

Here we concentrate on the process to compute the cross-sectional weights since it is the phase connected with the microsimulation model. This calibration phase is a stepwise procedure and it is carried out separately for each of the four longitudinal samples. First design weights are adjusted in order to account for non-response. Different non-response mechanisms (basically current non-response and non-response due to attrition) are considered. Then the weights are calibrated to obtain coherence with some external sources. In detail, constraints are built so that the weights reproduce a set of population counts of the cross-classification by gender, age and geographical area available from population registers.

Unfortunately some discrepancies, although not too prominent, may be detected between the estimation of aggregates of interest and data from other sources. In particular, it may happen to observe inconsistency between the estimates and the distribution by professional status available from the continuous Labour Force Survey (RCFL). This phenomenon could lead to possible biases in the estimation of aggregates and distributions of interest. The solution adopted was to insert an additional correction for non-response before the calibration phase to the population registers, calibrating to the information from the RCFL.

Final weights used for the computation of cross-sectional estimates for EU-SILC 2010 survey need to be updated in order to be utilized by the microsimulation model. If they are not adjusted the microsimulation model will provide results for the population at 2010 rather than for the target population, i.e. the population at 2012. To this aim the EU-SILC 2010 sample weights are calibrated using Istat demographic statistics updated at 2012 for the age and gender composition and 2012 RCFL data for the occupational status population distribution. The reason why this calibration step is chosen as adjusting step is for the minimum deviation from original weights property. In fact, original weights usually contain much valuable information on the relationship between sample and population. They are calibrated weights related to time *t* and, therefore, they reproduce the population characteristics used in the calibration process. In addition, original weights contain correction factors that take into account non-response.

In particular, less detailed constraints are defined with respect to the original set of constraints used in EU-SILC. In fact, a large number of constraints would lead to extreme variability of the weights and, consequently, not allow complete control on errors related to one or more key variables considered in the microsimulation model. Furthermore, an interesting problem is the double source of information on income (tax returns and sample, the latter derived from the responses of respondents). And it is reasonable to minimize the number of constraints related to demographic characteristics and employment status of the population, for example by giving up an excessive territorial disaggregation, to eventually extend the calibration also to known totals reported to the characteristics of the individuals presenting tax returns in all their possible forms (730 form, *Unico Persone Fisiche* form, 770 form for withholding agents).

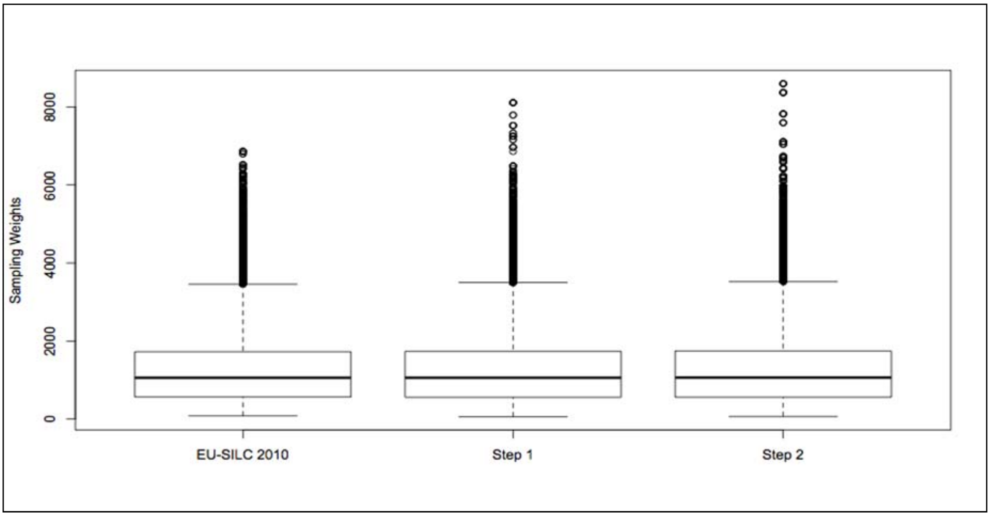
Therefore, to update EU-SILC 2010 sampling weights to the 2012 population distribution we proceeded in the following way:

- step 1: use the RCFL 2012 data to perform a phase analogous to the intermediate calibration stage in the original EU-SILC procedure;
- step 2: calibrate to the 2012 demographical information.

An important aspect to be analyzed when performing calibration is that the coefficient of variation (CV) of the final weights, i.e. the calibrated weights, should be comparable with the CV of the initial weights. In our case this is satisfied since the CVs for the original EU-SILC, step 1 and step 2 weights are respectively equal to 0.723, 0.738 and 0.735.

The following plots display some features of the EU-SILC sampling weights at the end of the calibration process to the 2012 external sources explained above.

Figure 2 - Distribution of the sampling weights for the EU-SILC 2010 survey: step 1 (calibration to RCFL 2012 survey) and step 2 (calibration to 2012 population registers)



Source: Elaborations on EU-SILC data

Figure 2 shows that the distribution of sampling weights is not substantially modified in the steps 1 and 2. In fact the first quartile Q1, the median and Q3 remain unchanged in the three distributions. The only difference is that the distributions related to steps 1 and 2 show larger extreme values with respect to the original EU-SILC sampling weights. But this difference seems to be unremarkable since the number of observations in steps 1 and 2 exceeding the maximum value of EU-SILC sampling weights is only 22 and 26 respectively, which is negligible if compared to the overall sample size (about 47,500 sampling units). Therefore, we can be confident that using the final calibrated weights in the microsimulation model will not create any significant deviation from the target population.

4. Social security contributions on employees and employers

4.1 Legislation, rules and modeling

In shaping the legislation ruling social security contributions in Italy, we have improved the treatment of some aspects, while maintaining the state of the art achieved by previous microsimulation models (ITAXMOD and MASTRICT). In particular, in ITAXMOD only the social contributions of employees were modeled in strict accordance with the rules, whilst in the case of the self-employed the contributions were computed applying an average of the statutory contribution rates pertaining to the main private professional pension funds. In MASTRICT, on the other hand, only the social security contributions charged onto employees were modeled, ignoring those paid by the employers. In building the social contributions module of FaMiMod, we have taken into account both the contributions payable by employees and those paid by the employers, considering in detail the statutory contribution rates of the different categories of self-employed workers.

In particular, in the social contributions module of FaMiMod, legislation in force in 2012 has been modeled taking account of the differences in the contribution rates relating to the individual characteristics of the worker: status (employees, contract workers, self-employed and professionals), professional qualification (production workers, clerical workers, junior managers, senior managers, apprentices), sector of activity (agriculture, manufacturing, construction, trade, information and communication, finance and insurance, public) and, for employees and contract workers, firm size (up to 15 employees, 16-49 employees, more than 50 employees).

Table 1 - Statutory contribution rates by type of worker: some examples – 2012

EMPLOYEES			
Employers and employees			
Rates by: sector, professional qualification, firm size, age and other characteristics			
APPRENTICES	employees	Employers	Employers
	rate if age is 15-29 years	rate	rate if firms has less than 9 employees
	5.84	10.0	0.0
SELF EMPLOYED			
Workers			
	rate	rate if age < 21 years	extra rate if income > 44.204 Euros
ARTISANS	21.30	18.30	0.01
SALES WORKERS	21.39	18.39	0.01
FARMERS	21.60	19.40	-
PROFESSIONALS			
	rate	extra rate if income > than ceiling	
ENGINEERS AND ARCHITECTS	13.5	0.03	
LAWYERS AND ATTORNEYS	13.0	0.03	
NOTARIES	33.0	-	
	rate	rate if age < 35 years	
LABOR CONSULTANTS	12.0	0.06	
…			
CONTRACT WORKERS (a)	rate	rates for retirees	
	27.72	18.0	

Source: INPS (National Social Security Institute) for employees, contract workers and self-employed; Private professional providers for each group of professionals.

(a) atypical workers.

Self-employed workers insured with INPS Funds (sales workers, artisans, farmers, tenant farmers and sharecroppers) were considered separately from the professionals, who are insured with specific private professional pension funds (lawyers, notaries, accountants, etc.). In addition, the rules on upper ceilings and minimum statutory rates and – as far as possible – special rates applied to particular conditions (e.g., retirees) have been taken into account. Table 1 presents some examples of the main statutory rates applied to different types of workers. The identification of professional qualification, sector, firm size and profession has been carried out on the basis of the information available in the EU-SILC Survey.

Table 2 - Social contribution rates, revenue and tax base: comparison among EU-SILC Survey aggregates (2009), FaMiMod model estimates (updated to 2012), and National Accounts statistics (2009 and 2012) *(millions of euros)*

	EU-SILC - 2009	National Accounts - 2009	FaMiMod - 2012(a)	National Accounts - 2012
EMPLOYEES				
Employers (b)				
average aggregated rate (%)	25.24	24.87	26.20	24.90
revenue	160,000	161,971	173,700	167,459
income from employment	634,000	651,354	662,000	672,594
Workers				
average aggregated rate (%)	8.82	8.51	8.69	8.23
revenue	41,800	40,527	45,600	40,381
gross wages	474,000	476,179	523,000	490,827
CONTRACT WORKERS (c)				
Employers				
average aggregated rate (%)	8.14		11.64	
revenue	1,620		2,620	
income from employment	19,900		22,500	
Workers				
average aggregated rate (%)	4.64		6.36	
revenue	812		1,310	
gross wages	17,500		19,800	
SELF EMPLOYED				
Workers				
average aggregated rate (%)	14.84		19.69	
revenue	27,600		25,400	
income from employment	186,000		129,000	

Sources: National Accounts, Annual sector accounts, Sequence of accounts by institutional sector, Total economy, Resources, years 2009 and 2012, (Edition: October 2013; aggregates: compensation of employees, wages and salaries, employers' actual social contributions, employees' social contributions)

European Union Statistics on Income and Living Conditions (EU-SILC) 2010 for Italy, 2009 incomes

Istat: Microsimulation model of household taxes and benefits (FaMiMod), 2009 incomes updated to 2012

a) we used the survey weight for EU-SILC 2009 data while for the FaMiMod model we used the reweighting updated to 2012

b) the employers contributions for employees in FaMiMod 2012 include severance pay (TFR)

c) atypical workers.

The accuracy in modeling the legislation allows greater precision in calculating the effects of hypothetical changes of the rules of social contributions, through a more detailed design of the simulation exercises. Table 2 shows the average aggregated rates, the average of individual rates, the total revenue and the tax base of the social security contributions, providing a comparison between the results of the EU-SILC Survey, estimates of the FaMiMod model and related National Accounts statistics[11]. As can be seen from Table 2, the model estimates for contribution rates, revenue, and tax base for the employees are in line with those provided by the National Accounts. For the self-employed the comparison is only between the results of the EU-SILC Survey and the estimates of the FaMiMod model. There are many differences in the definitions of the aggregates and asymmetries between the composition of the aggregates of the survey and those of the model if compared to those of National Accounts (for example, contract workers are considered together with the self-employed and are not easily separable from the latter).

4.2 Labour cost, wages and tax wedge

Using the FaMiMod model, the effects of economic and fiscal policy measures on the tax wedge can be calculated. The tax wedge equals the difference between the labour cost sustained by the employer and the net wage received by the employee. The tax wedge corresponds to the levy, in terms of taxation on income and of social contributions made by the State. One can distinguish between the direct tax part of the tax wedge on labour income and that which relates to social security paid by employers and workers. The whole burden of the levy on labour is paid by different subjects, the employer and the employee and, as we saw in the previous section, the contribution rates are very different, even if considering only employees, depending on the type of work, qualification, sector, firm size.

As a consequence, changes in the components of contributions may, for example, have different effects on labour costs of employees who work in different areas or have different qualifications. The components of labour costs have been estimated with FaMiMod, using the EU-SILC Survey data of 2009, updated to 2012. The calculation of the components of the tax wedge was made separately for the income of employees and contract workers and, in the part relating to the social contributions, both those paid by workers and those paid by the employers were estimated[12].

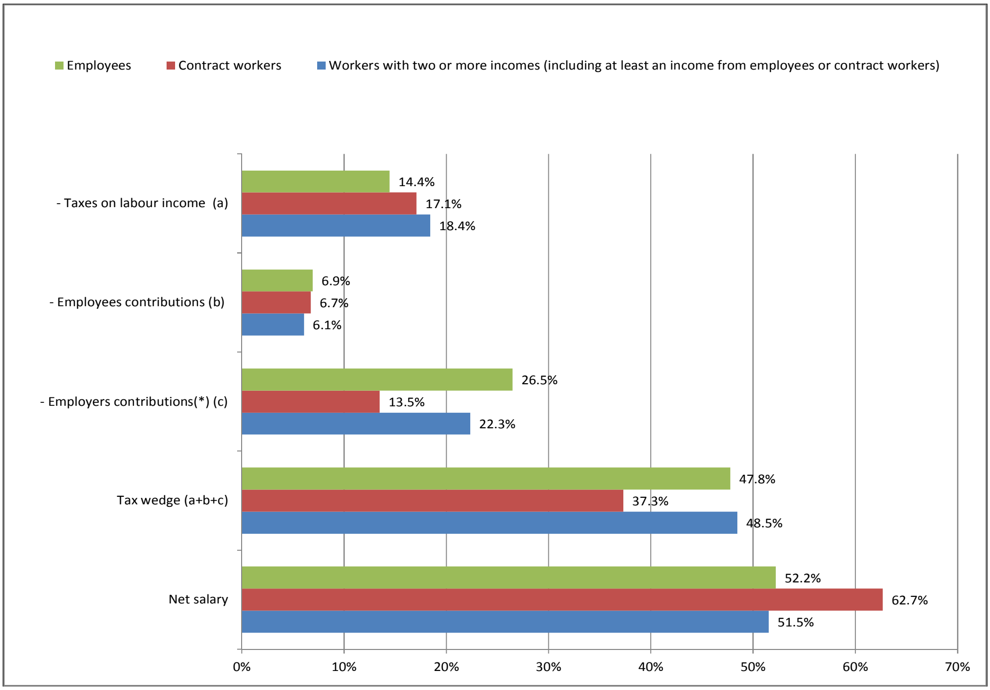
In the graphs below, workers are distinguished depending on whether they receive only one income (employees or contract workers) or two or more incomes (including at least one as employees or contract workers). In the case of two or more incomes, generally, the tax wedge on labour income and employees or contract workers also depends on the presence of self-employment income[13].

[11] The National accounts data are referred to 2009 for comparison with EU-SILC Survey data and to 2012 for comparison with FaMiMod model estimates updated to 2012.

[12] The tax wedge presented in this section has been calculated as the ratio between the total revenue from taxes and contributions on labour income and the total labour cost. The incidence of the tax wedge and its various components is calculated as the ratio between the revenue and the total labour cost. For algebraic reasons, this estimate does not necessarily coincide with the average of the individual effects, which is very much influenced by extreme values. In the case of more than one income, the tax wedge on employees and contract workers may also depend on the simultaneous presence of other income from self-employment. The income of self-employment is not considered in the calculation of the wedge. However, the amount contributes to determining the effective tax rate paid on the total labour income, the model then applies it 'pro-rata' to the incomes of employees and contract workers.

[13] It is considered that you might have a self-employment income even if it is not included in the calculation of the wedge. In the case of more than one income, the tax wedge on employees and contract workers may also depend on the simultaneous presence of other income from self-employment. The income of self-employment is not considered in the

Figure 3 - Wages, tax wedge and its components *(in % of labour cost)*



Source: Istat, Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012
(*) The employers contributions for employees in FaMiMod includes severance pay (TFR).

Wages, the tax wedge and its components in relation to total labour cost are presented in Figure 3, depending on whether workers receive one or two or more incomes. In 2012, the recipients of an employee income receive, on average, a net wage which is slightly more than half of the average labour costs (52.2%). The average value of the total tax wedge for employees amounted to 47.8% of the labour cost from which one can distinguish the component of social contributions (33.4%) and that of taxes on labour (14.4%). Social security contributions are, in general, the highest tax wedge component and are made of a more significant share charged to the employer (26.5%), and a less costly share paid by the worker (6.9%). However, as workers are affording taxes on labour income along with social contributions, together they account for 21.4% of labour cost. The recipients of a contract worker income, compared to that of pure employees, have an average net wage which represents over 60% of the average labour cost. The tax wedge is approximately 37.3% of the labour cost, of which 20.2% is attributable to contributions and 17.1% to taxes on labour income[14]. Moreover, contributions payable by the employer are 13.5% and those

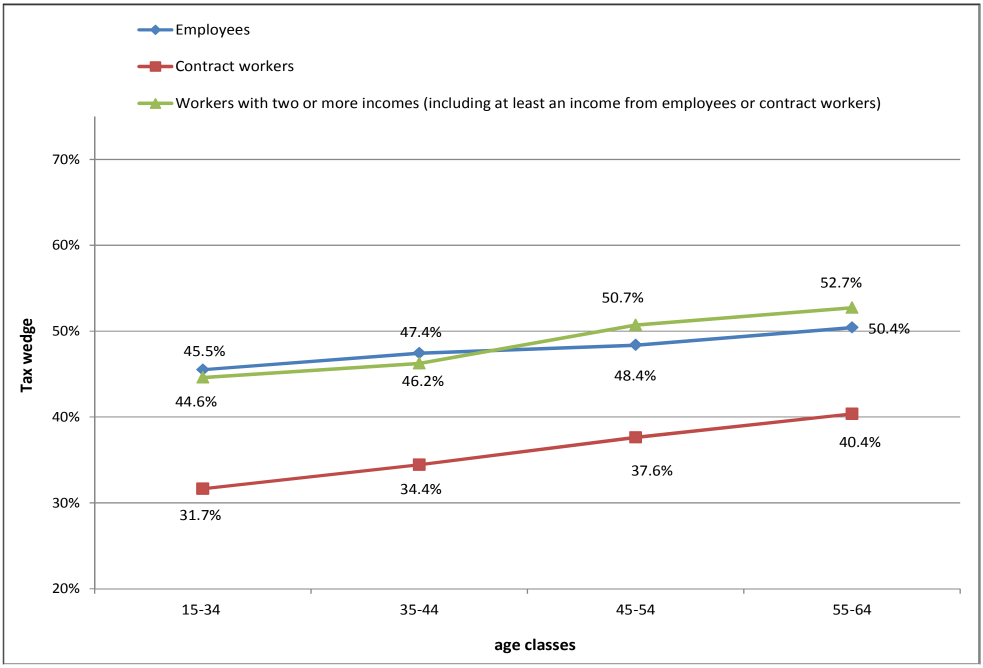
calculation of the wedge. However, the amount contributes to determining the effective tax rate paid on the total labour income, the model then divides them 'pro-rata' on the incomes of employees and contract workers.

[14] Contract workers have a lower statutory contribution rates than employees (the statutory contribution rate is 27,7 versus 33% in 2012) and also they do not have the severance pay (TFR) that is instead included in the employers contributions of the employees.

paid by workers are 6.7%. The earners of two or more income have a net wage that is approximately 51.5% of the labour cost. The tax wedge is 48.5% of the labour costs, with social contributions accounting for 28.4% while taxes total 18.4%. The share of social contributions paid by employers is 22.3%, and the workers charges are 6.1%.

Among the factors influencing contributions and the taxation of labour income there are, in particular, the age and gender structure of the population. These two demographic characteristics affect the level of employment, the rules of retirement and, in general, the social protection system. Obviously, the higher the cost resulting from the system of social protection, the greater the social contributions charged to employers and employees.

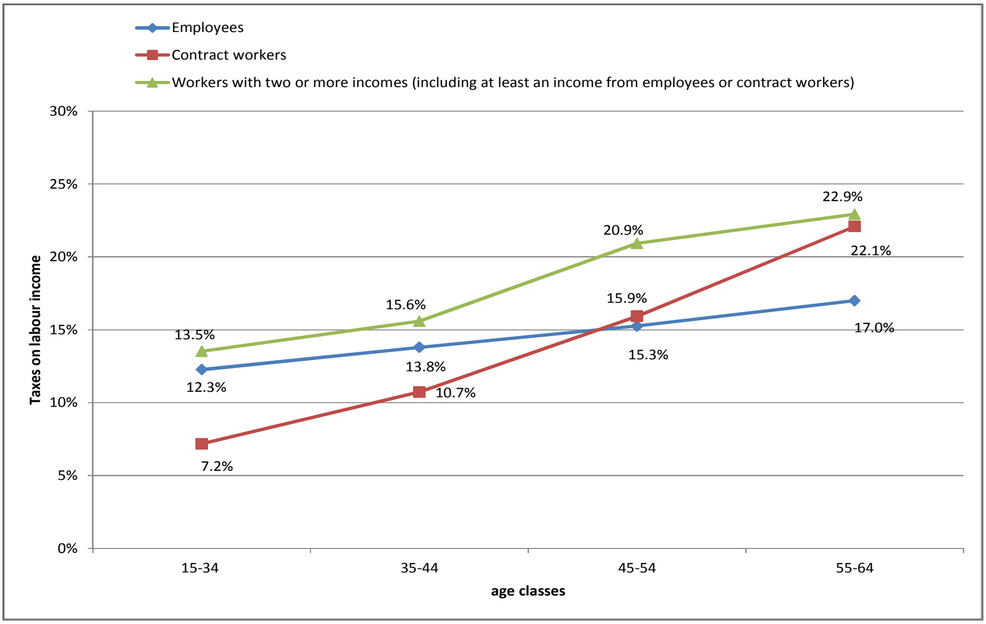
Figure 4 - Tax wedge by age *(in % of labour cost)*



Source: Istat, Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012

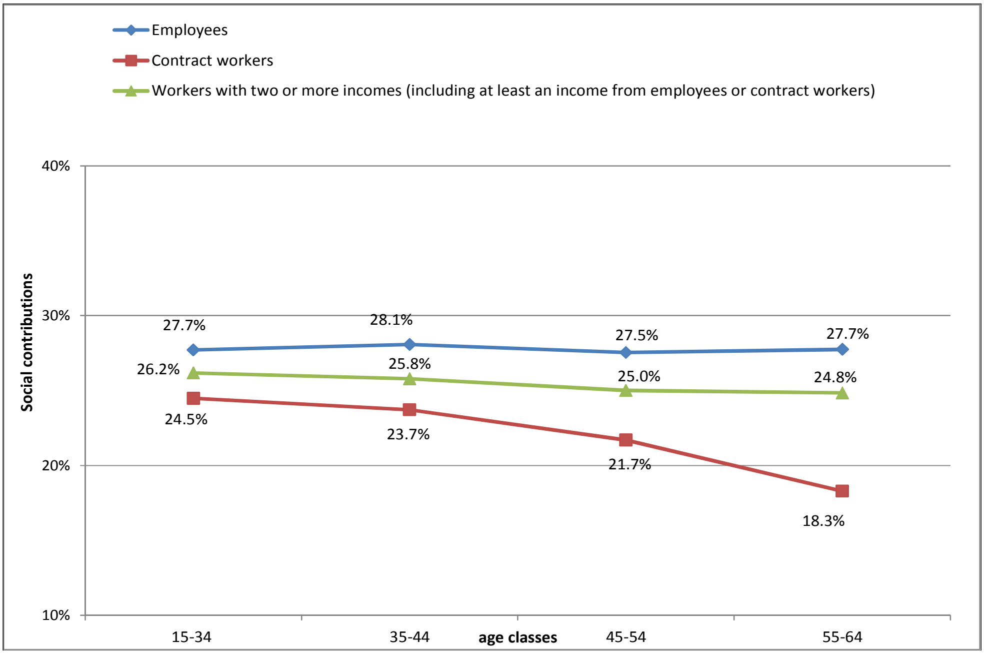
In Figure 4 the tax wedge for the earners of one or more income during the active life cycle is presented, distinguishing by age groups. The tax wedge and its component of labour taxes increase with age, as expected, following the upward trend of income during the active life (Figures 5 and 6). This trend is primarily due to the taxation of labour income, to a greater extent for the contract workers, where the figure grows from 7.2%, for workers less than 35 years old, to 22.1%, for those aged 55-64. The social contributions component, however, slightly decreases along the life cycle for all ages and for all categories of workers observed, to a greater extent for contract workers (see Figure 6).

Figure 5 - Taxes on labour income by age classes *(in % of labour cost)*



Source: Istat, Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012

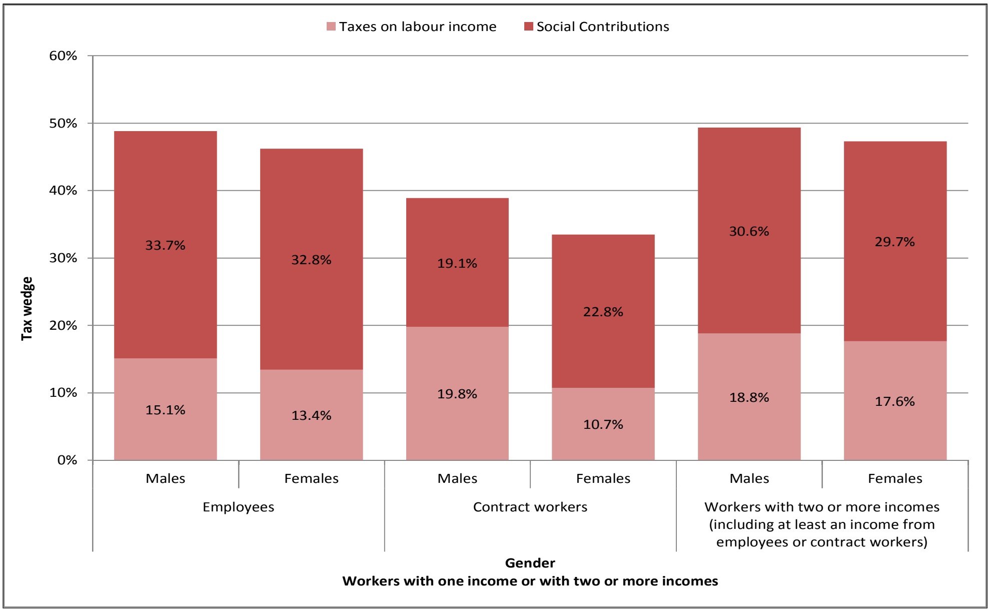
Figure 6- Social contributions by age classes *(in % of labour cost)*



Source: Istat, Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012

Figure 7 shows the gender breakdown of the tax wedge. Women have a lower overall tax burden than men (about 3 percentage points lower for employees, almost 5 for contract workers) and this trend is also observed in the tax on labour component of the tax wedge. The only exception is among contract workers, and it regards social contributions of women. Women showing an average rate higher than that of men also have an average value of labour cost much lower (about half) than that of men (50% of men have labour costs lower than 20,000 euros while 50% of women have labour costs under 10,000 euros). This is probably because women are more likely to be subject to the higher rates determined by the minimum rate of contribution specified by law.

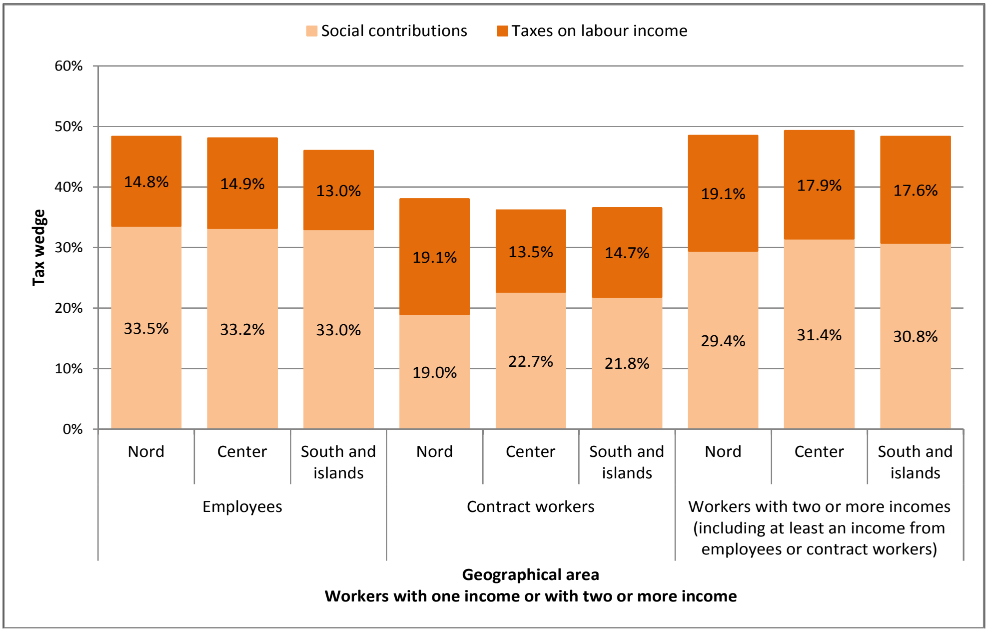
Figure 7 - Tax wedge by gender *(in % of labour cost)*



Source: Istat, Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012

The tax wedge is then disaggregated by geographical area (Figure 8). It is slightly higher in the Center and in the North than in the South and the Islands, for all the considered workers and for the two components of the wedge (taxes and contributions). The employees, along with workers with two or more incomes - including at least an income from working as an employee or as a contract worker - do not show significant differences among geographical areas. The variation in the tax wedge among areas range between 48.3% (in the North) and 46% (in the South and in the Islands) of the labour cost for employees and between 38% (North) and 36.2% (Center) for contract workers, settling, however, at around 49% for workers with two or more incomes in the three main areas. Some differences among the contract workers are found in the components of the tax wedge. On average, the lower values of the labour taxes are attributable to the workers who live in the South and the islands (13% of labour costs versus 14.8% in the North and 13.5% in the Center), while the lowest incidence of contributions is found among those living in the North (19% compared to 22.7% in the Center and 21.8% in the South and the islands).

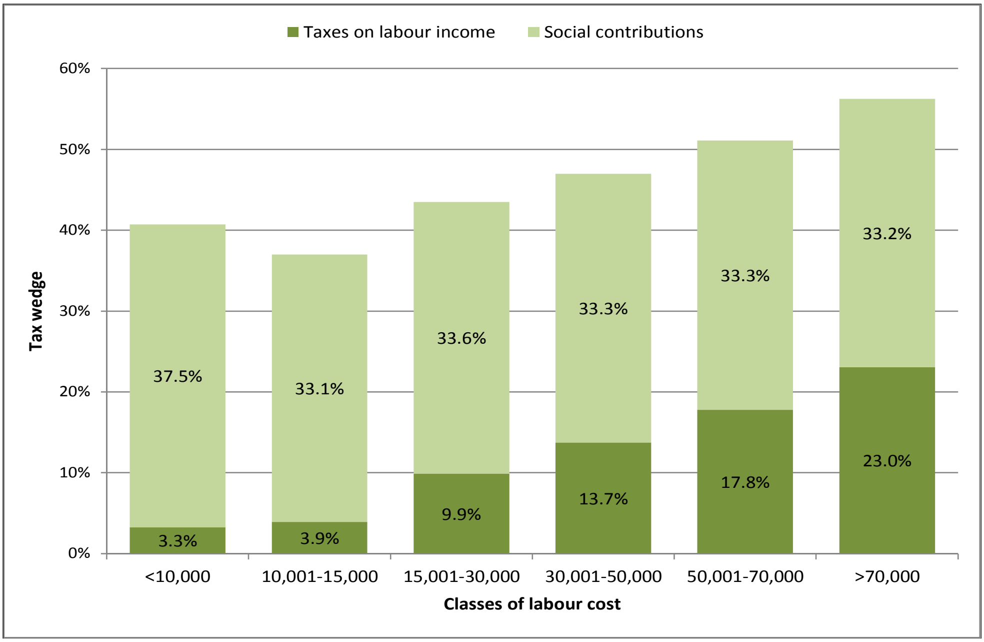
Figure 8 - Tax wedge by geographical area *(in % of labour cost)*



Source: Istat, Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012

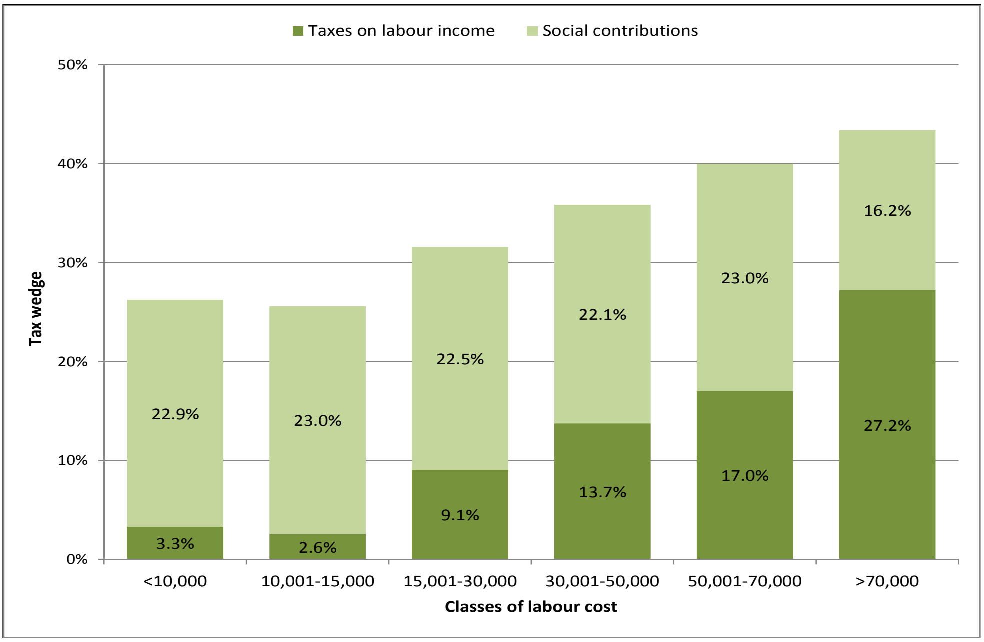
The distribution of taxes and contributions is observed looking to the tax wedge as a percentage of the labour cost. For all types of workers considered the wedge widens when labour costs increase, as expected, driven by the portion of the wedge represented by progressive taxes on labour income (Figures 9, 10 and 11). The exception to this trend is in the first class of labour cost where the wedge is higher than that of the next class. Due to the minimum rate of contribution specified by law, for an important part of workers whose labour cost is less than 10,000 euros, the tax wedge is 40.7%, higher than that of workers whose labour cost falls between 10,000 and 15,000 euros, that is 37%. In the next classes of labour costs, the wedge starts to increase slightly. The rising wedge is explained by the increase in labour taxes that increases the labour costs while contributions are reduced, decreasing from 37.5% of labour costs below 10,000 euros to 33.2% for those greater than 70,000 euros. In the class of higher labour cost, due to the effect of the ceilings, the contributions are slightly lower, whereas in the class under 10,000 euros, due to the minimum rate, they are higher. This trend can be seen for the other two categories of workers: the contributions component decreases with labour costs, while taxes on labour increase. The decline of the wedge in the transition from the class of labour costs lower than 10,000 euros to that between 10,000 and 15,000 euros also occurs for the other workers considered: for contract workers the wedge goes down slightly from 26.2% to 25.6%, while for workers with two or more incomes the wedge is reduced to an extent similar to that of employees, declining from 40.2% to 36%.

Figure 9 - Tax wedge by labour cost – Employees *(in % of labour cost)*



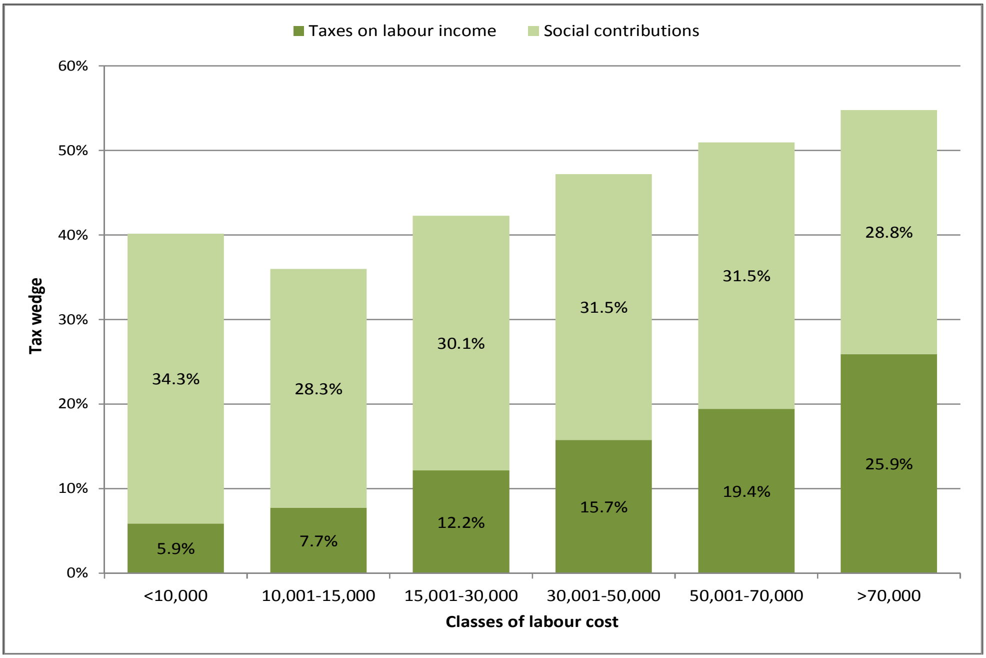
Source: Istat, Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012

Figure 10 - Tax wedge by labour cost – Contract workers *(in % of labour cost)*



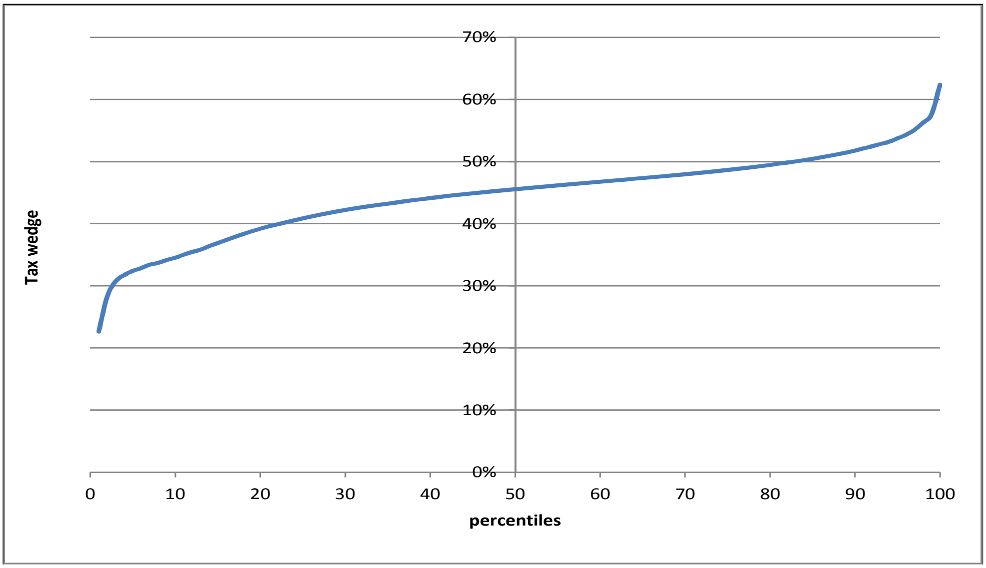
Source: Istat, Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012

Figure 11 - Tax wedge by labour cost – Two or more income earners *(in % of labour cost)*



Source: Istat, Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012

Figure 12 - Tax wedge – Employees *(average individual rate)*



Source: Istat, Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012
(a) The tax wedge is computed at the average individual level.

Finally, in Figure 12, the distribution of the tax wedge[15] shows that 50% of workers with one labour income as an employee has a tax wedge which accounts for slightly less than 47% of the total labour cost.

As we have seen the tax wedge in Italy is particularly high. In OECD (2013) it is shown that, in 2012, the percentage incidence of taxation on labour income and of social contributions on labour cost[16] in Italy amounted to 47.6%, placing Italy in the sixth position among the 34 OECD countries, after Belgium, Germany, France, Hungary and Austria (Table 3). However, the level of the tax wedge in each country reflects the system of taxation and redistribution that each country has adopted, which makes it difficult to evaluate the effects of measures to intervene on the tax wedge.

The reduction of the tax wedge is one of the objectives of the Budget law ("Legge di stabilità") for 2014[17] for two reasons related to the short-term crisis and the long-term lack of economic growth. The first reason is to help increase productivity through greater competitiveness of firms and thus a potential restart of the economic growth: for newly hired workers with permanent contracts made in 2014 to increase the employment base, there is a deduction from the Regional Tax on Productive Activities (Irap) up to 15,000 euros for a three-year period, resulting in a reduction of the tax due by the employer[18]. The second reason is to achieve the objectives of redistribution with respect to certain categories of workers by increasing tax credits for earned labour income, thus reducing the tax burden and therefore increasing the net wage of the employee.

[15] The distribution displays the dispersion of the tax wedge computed at the average individual level, whilst in the previous graphs the tax wedge was equal to the ratio between total revenue and total labour cost.

[16] The percentage of labour cost is taken with reference to a single individual without children who earns the income of the average production worker (see OECD, 2013).

[17] "Disposizioni per la formazione del bilancio annuale e pluriennale dello Stato", Law n. 147/2013, art. 1, paragraph 127.

[18] The calculation of the tax wedge based on FaMiMod model does not encompass the Irap component paid by the employer. However, the model provides an estimate of the Irap relating to the personal income of the sampled self-employed workers.

Table 3 - The components of the tax wedge as a percent of labour costs in 2012[(a)]

Country[(b)]	Income tax	Employee SSC	Employer SSC[©]	Total tax w edge[(d)]
Belgium	22.1	10.8	23.2	56.0
France	10.2	9.5	30.6	50.2
Germany	16.0	17.3	16.4	49.7
Hungary	12.8	14.4	22.2	49.4
Austria	12.3	14.0	22.6	48.9
Italy	16.1	7.2	24.3	47.6
Sw eden	13.6	5.3	23.9	42.8
Finland	17.7	6.2	18.6	42.5
Czech Republic	8.8	8.2	25.4	42.4
Slovenia	9.4	19.0	13.9	42.3
Greece	6.9	12.8	22.2	41.9
Spain	13.5	4.9	23.0	41.4
Estonia	12.7	2.1	25.6	40.4
Slovak Republic	7.4	10.5	21.8	39.6
Netherlands	14.9	13.9	9.7	38.6
Denmark	36.2	2.7	0.0	38.6
Turkey	11.1	12.9	14.2	38.2
Norw ay	19.1	6.9	11.6	37.6
Portugal	8.7	8.9	19.2	36.7
Luxembourg	13.8	11.0	11.0	35.8
OECD	13.1	8.2	14.4	35.6
Poland	5.8	15.3	14.4	35.5
Iceland	26.8	0.4	7.2	34.5
United Kingdom	14.0	8.5	9.8	32.3
Japan	6.6	12.0	12.6	31.2
Canada	13.6	6.6	10.6	30.8
United States	15.6	5.1	8.9	29.6
Australia	21.6	0.0	5.6	27.2
Ireland	13.4	2.9	9.7	25.9
Sw itzerland	9.7	5.9	5.9	21.5
Korea	4.4	7.4	9.2	21.0
Israel	7.5	7.3	4.4	19.2
Mexico	7.3	1.2	10.5	19.0
New Zealand	16.4	0.0	0.0	16.4
Chile	0.0	7.0	0.0	7.0

Sources: OECD (2013); country submissions, OECD (2012)

(a) Single individual without children at the income level of the average worker

(b) Countries ranked by decreasing labour costs

(c) Includes payroll taxes where applicable.

(d) Due to rounding, the total tax wedge may differ by one or more percentage points from the sum of the components. For Denmark, the Green Check (cash benefit) contributes to the difference as it is not included in the components.

5. The Personal Income Tax

5.1 Taxable incomes, gross and net tax liabilities

The Italian personal income tax (Irpef - *“Imposta sul reddito delle persone fisiche”*) applies increasing tax rates to the income, net of tax deductions:

$$T^G = \tau_i(Y^{TN} - L_i) + \Delta Y_i \qquad \text{if } L_i < Y^{TN} <= L_{i+1}$$

$$\text{where } \Delta Y_i = \Delta Y_{i-1} + \tau_i(L_{i+1} - L_i) \text{ and } \Delta Y_1 = 0$$

In the above formula, T^G is the gross tax liability for a taxable income Y^{TN} (net of tax deductions) and L_i its lower bound of the *i*-th income bracket (*i*= 1,…,5). The final tax liability is obtained by subtracting the tax credits from the gross tax.

Some incomes are totally tax exempt. The most important examples are the social security public transfers for low-income elderly (*“Assegno sociale”* and *“Pensione sociale”*), for children (*Family allowances*, *Maternity allowances*), for disabled persons.

Taxable income Y^{TN} is net of three main deductions:

- Mandatory social security contributions on primary (market) incomes
- Incomes taxed separately at source (*e.g.* interests on financial assets)
- Deductible expenses (*e.g.* voluntary contributions to private pensions plans)

Tax evasion is equivalent to another deduction from the tax base and must be estimated, even though approximately, to ensure consistency with the aggregate revenues recorded by the tax administration and encompassed in the National Accounts.

In pre-existing microsimulation models, correction coefficients were applied to components of total income showing a severe misalignment in their weighted sample total with respect to tax aggregates, in particular income from self-employment.

In the Italian edition of the EU-SILC project, when both the administrative files and the survey report it, income from self-employment is set equal to the maximum value between individual (anonymous) information on: (i) the net self-employment income resulting from the tax return and: (ii) the net self-employment income reported in the survey questionnaire.

This procedure is adopted to minimise either under-estimation due to tax evasion in the administrative data or under-reporting in the survey data, depending on which of the two is larger[19]. Among the individuals for which both sources contain self-employment incomes, the record linkage reveals that under-estimation is more frequently observed in the tax data than in the survey data. It turns out, moreover, that self-employment income in the integrated dataset is more unequally distributed than in the survey.

Tax exempt income of any kind, including tax avoidance, is encompassed in the

[19] The procedure requires the exact matching of survey with administrative data. With respect to the exclusive use of survey data, the record linkage increases substantially the number of percipients and the average self-employment income, resulting in aggregate estimates that are closer to the National Accounts figures.

disposable income of the individuals and their households and should be accounted for in the net income simulated by FaMiMod. In fact, when self-reported income is greater than the corresponding administrative value, the model uses such a difference at the micro level as a proxy for the amount of income that is not reported to the tax authorities. No correction is made when the administrative income is greater than that reported in the survey. The ratio of unreported to net income in the base year is then multiplied by the updated self-employment income in subsequent years to estimate the amount undeclared to the tax agency. As yet, the impact of the correction is relevant, as can be seen from Table 4.

Table 4 - Composition of overall income for tax purposes: FaMiMod versus tax returns *(thousands of euros)*

	Tax Returns 2012 (2011 incomes)	FaMiMod (2012 incomes) without corr_AUT	with corr_AUT
	(a)	(b)	(c)
Employee income	**422,904,039**	**467,462,666**	**467,462,666**
Pensions	**233,863,552**	**245,003,631**	**245,003,631**
Land	**1,335,021**	**2,081,063**	**2,081,063**
Buildings	**35,000,761**	**35,279,919**	**35,279,919**
Enterprise	***31,826,979***		
Self-employment	***33,906,366***		
Partnership	***35,892,693***		
Other income	***7,307,435***		
Total self-employment income	**108,933,473**	**205,980,020**	**128,267,218**
Optional separate taxation	517,190		
Maternity leave		383,018	383,018
Overall income	**800,293,855**	**956,190,317**	**878,477,515**

Source: (a) http://www.finanze.gov.it/stat_dbNew/index.php
(b) and (c): simulations with Istat Microsimulation model of household taxes and benefits (FaMiMod; corr_AUT: correction of self-employment income). To assure homogeneity with (a), Buildings and Overall income include imputed income of owner-occupied housing, following the tax law in force until 2011.

These data coming from different sources allow an analysis of the relationship between under-reporting and tax evasion (see Di Marco, 2007).

The model can also provide an estimate of the share of Irap ("*Imposta Regionale sulle Attività Produttive*") relating to self-employment incomes. Depending on the scope of the analysis, this charge on companies can be optionally included in the total tax liabilities of the individuals. Indeed, Irap is not an income tax *per se*, the value added being its tax base. However, for the share of value added that consists of compensations for the work of the self-employed who own the business, Irap can be viewed as an additional tax on income. Moreover, this assures consistency with EU-SILC target variable for gross income from self-employment used as input, that includes a share of Irap estimated by applying the statutory tax rate to the share of self-employment in total taxable income

The sum of the tax credits determines algebraically the upper bound of the *no tax area*, below which no tax is due. If the sum of the tax credits exceeds gross tax liabilities, the difference is lost, since as a general rule there is no negative income tax[20]. A part of the fiscal benefits can thus be canceled for the taxpayers with the lowest incomes, namely those included in the *no tax area*.

5.2 The tax credit module

For the simulation of tax credits for employment, similar incomes and pensions the EU-SILC source supplies the basic information, with the exception of months worked by contract workers (a category of self-employed who are treated like employees under this respect). When missing, these had to be estimated as briefly sketched above (see section 2). On the other hand, unlike the Bank of Italy SHIW, months of income are present for various kind of unemployment benefits that are treated as employee income with respect to tax credits.

For the simulation of tax credits for dependent relatives (spouse, children, and other), pre-existing microsimulation models based on SHIW had to reconstruct the relevant tax unit as a preliminary, splitting multinuclear households when each nucleus included potential beneficiaries and/or restructuring the household in order to set a potential beneficiary rather than a dependant at the head of the unit. The EU-SILC source allows to skip this stage, since it includes information on nuclear families within the household (defined by parental and couple relations) that can be viewed as a suitable proxy for the tax units entitled to tax credits (as well as to family allowances, see section 6).

The modelling of tax credits takes into account all the important details of the tax law, already considered by the most advanced pre-existing models. In particular, the model considers the possible trade-offs in the choice of the most convenient tax credit for the first child of a lone parent and, moreover, in the attribution of the full tax credit for the dependent children to the highest income parent as an alternative to the fifty-fifty splitting between both parents when none of them is dependant. For lone parents, the tax credit for the first dependent child could equal, if more beneficial to the taxpayer, the tax credit for the spouse of an ordinary household. Actually, these tax credits share the same amount (in 2012) in correspondence of very low income levels. The first child tax credit, however, is constant for a broad income bracket, whilst the spouse tax credit is more strictly related to the taxpayer’s income.

For couples with children, the choice between a full tax credit to the highest income parent and a splitting of it in two halves only arises when the parent with the lowest income cannot benefit (totally or partially) of the tax credits he is entitled to (in Italian, “*incapiente*”). In fact, with few exceptions, the Italian tax schedule does not entail negative income taxes, so that in this case the highest income parent may be entitled to a greater effective benefit, even though his tax credit decreases as income grows.

Minimum values of tax credits provided for low income employees (including contract workers) and pensioners, aimed at limiting tax credit cuts due to a limited number of months of activity, have also been modelled. This detail proves particularly useful in order to check simulated values against individual (anonymous) tax record data associated with the sample, because in this case values are fixed rather than highly sensitive to small

[20] The special tax credit for households with three or more dependent children, however, is an exception, as are tax credits for rents (see below).

variation of income as is the rule, due to the linear decreasing mechanism introduced in 2002 when tax credits were turned into tax allowances and confirmed in 2007 when tax credits were restored.

When simulating tax credits, some items are particularly subject to overestimation, that is the tax credit for other dependent relatives and the (payable) tax credit for rents. In the first instance, entitled taxpayers do not always claim this benefit (that is, take-up is less than 100%); on the other hand, in some cases the model is not able to assure that the family relationship between potential claimant and dependant complies with the rules set by the tax law, aimed at excluding less close relationships. In the second instance, the main source of overestimation is the phenomenon of undeclared rents, since a registered rental agreement must be referred to when claiming for this tax credit.

Table 5 - From overall income for tax purposes to net tax: FaMiMod versus tax returns *(thousand of euros)*

	Tax Returns 2012	FaMiMod (2012 incomes)	
	(2011 incomes)	with sel_DETR	without sel_DETR
	(a)	(b)	(c)
Overall income	800,293,855	878,477,515	
Allowance for owner-occupied housing	8,510,433	8,599,632	
Tax allowances	22,400,416	36,454,276	
Taxable income	772,219,281	834,435,128	
Gross tax	208,215,753	223,712,968	
Tax credits for dependent relatives	11,289,654	12,237,558	
of which: other relatives	*230,862*	*285,670*	*438,703*
Tax credit for employment and pensions	41,467,876	42,185,390	
Tax credit for medical and other expenses	5,476,874	5,838,837	
Tax credit for home improv. & maintenance	2,457,789	2,175,808	
Tax credit for rents	146,926	187,935	*541,105*
Other tax credits	1,318,279	1,156,277	
Total tax credits	62,112,973	63,784,205	
Net tax	152,219,369	164,853,295	

Source: (a) http://www.finanze.gov.it/stat_dbNew/index.php (tax credit for other relatives: tax records associated with the sample)

(b) and (c): simulations with Istat Microsimulation model of household taxes and benefits (FaMiMod; sel_DETR: correction of selected tax credits). To assure homogeneity with (a), Overall income includes imputed income of owner-occupied housing and the corresponding allowance has been calculated, following the tax law in force until 2011.

In both cases, the chosen solution (already experimented with MASTRICT model) is the application of a random selector. This allows the exclusion of a number of beneficiaries, limiting the overestimation compared to aggregate tax record data (Table 5). The specific innovation in FaMiMod is to use the individual (anonymous) tax record data as an additional filter, in order to prevent real beneficiaries of this tax credit from having simulated values set to zero as a result of the random process.

As stated above (see section 2), tax credits based on deductible charges (health expenses etc.) which cannot be simulated on the basis of sample information are imputed using individual (anonymous) information stemming from tax records linked to sample units. For the residual fiscal item "other tax credits", care is taken to minimise possible duplications with respect to simulated tax credits, in particular the tax credit for rents. Since statutory values show a limited differentiation, it is possible to detect cases where the "other tax credit" might correspond in fact to a tax credit for rents: in these cases, the value to be imputed from tax records is set to zero when simulated tax credit for rents exist.

In this first release of FaMiMod, the beneficiaries of alimony could not be included among those entitled to tax credits for employment, similar incomes and pensions, because information on alimony is missing from the 2010 release of EU-SILC taken as a source. In the next release the information should be restored, thus allowing to plug the gap.

In perspective, further developments of the tax credit module are possible by further exploiting the potential of the database. For instance, individual tax information linked to sample units allows to track individuals with dependent relatives living outside the household (for instance because they are divorced), up to now invisible to microsimulation models. The same source of data can help improving the correspondence between EU-SILC nuclei and tax units, particularly for components that are marked as "isolated" rather than children only because of their marital status (for instance, widows): these should be traced back to a nucleus when they can give rise to a tax credit.

As regards disability, a well-known gap in sample surveys on incomes, some information would be useful in order to include in the simulation the additional tax credit for disabled children. As a first approximation, proxies of this condition developed in other modules of the model (for instance, family allowances), however incomplete, could be exploited.

5.3 Regional and Municipal Income Taxes

The establishment of additional regional and municipal taxes comes as part of the process of fiscal decentralization of the state, in order to implement fiscal federalism[21].

The additional regional and municipal taxes are direct taxes that apply in addition to the personal income tax (Irpef) and are paid to regions and municipalities. The tax base on which both additional taxes are calculated is the same, and it consists in the sum of all personal incomes, net of deductible expenses. Nevertheless, a taxpayer is subject to additional taxes only if his personal income tax is due. The rates established by the region and the municipality of residence, within the limits of the maximum rate, must then be applied to the tax base. Additional regional and municipal tax rates are divided into two components: a partnership tax rate, which is compulsory and established by the central

[21] The regional additional tax was established by Legislative Decree n. 446/97 (art. 50) with effect from 1 January 1998. The municipal additional tax was established by Legislative Decree n. 360/1998 (Art. 1), and has been applied since 1999.

government[22] and an optional tax rate that can be introduced at the local level, within the limits of national law. The authorities of both regions and municipalities have legislative power over the determination of tax liability, as well as over any exemption based on taxpayers characteristics.

The additional municipal tax is ruled by the municipality[23] which may also introduce a threshold for exemption from tax for certain types of income (pension or employee or self-employed) and/or for the family composition, in relation to specific income thresholds. The system of local taxes that is currently in force has a wide heterogeneity in the application of different methods of tax levy, due to the decentralization at the regional and municipal level of the jurisdiction that defines the parameters relevant for this tax. In an attempt to simplify some local authorities adopt a single rate, while others have implemented a system of different rates. Municipalities may establish a variety of differential rates: however, they should articulate them according to the brackets of the personal income tax (*Irpef*) nationwide.

In the FaMiMod model there are two modules for additional taxes, one dedicated to the additional regional tax and the other to the municipal one. The module on the additional regional tax takes into account any tax exemptions or tax rates and increases for the entirety of the Italian regions.

As to the module on the municipal tax, information provided by individual municipalities has required a good deal of interpretation and systematization. All the information available in 2012 for the calculation of rates, exemptions and facilities has been taken into account to model the tax. The modeling, however, only concerns the municipalities represented in the EU-SILC survey (just over 10% of Italian municipalities), which are not necessarily representative of the distribution of the average rates of local taxes in Italian municipalities.

To assess the progressivity of additional regional tax structure, in Figures 13 a, b, c and d we consider the average rate[24] according to the different personal income tax brackets (Irpef) by region in 2012. Most regions (Valle d'Aosta, Trentino, Veneto, Friuli Venezia Giulia, Lazio, Molise, Campania, Basilicata, Calabria, Sicily, Sardinia) essentially have a proportional single tax rate, with the exception of the lowest income bracket (under 15,000 euros) where the incidence of low-income taxpayers located in the no-tax area results in a lower average tax rate. Other regions (Liguria, Emilia Romagna and Piedmont) have a lower tax rate in the two lowest income brackets (up to 28,000 euros), while two regions (Tuscany and Umbria) have a single tax rate in central income brackets, a lower tax rate for those on lower incomes (under 15,000 euros) and an higher one for higher incomes (over 75,000 euros).

[22] The basic rate for the additional regional tax is set by Central Government to 1.23%, while it was 0.9% previously. Regions can apply variations up to 0.5% on this rate in 2012, that is the limit established by national law (DL n. 138/2011 converted into Law no. 148/2011). In addition, regions with a budget deficit in health care have the option of applying the maximum rate of 2.03% (Legislative Decree n. 68/2011, implementing decree on fiscal federalism).

[23] The rate, which can vary between municipalities cannot exceed 0.80%, as expressly provided by law (Law no. 296/2006 art.1, paragraph 142). Previously, the optional additional tax rate could not exceed the maximum limit of 0.5%, with an annual increase of no more than 0.2%. The municipalities may decide to change the rate by 31 December of each year (Art. 28, Law no. 342/2000).

[24] The average tax rate is calculated as the ratio between the revenue from the additional regional tax and the corresponding tax base.

Finally, in some other regions (Marche, Lombardy, Alto Adige, Abruzzo and Puglia) there is a progressive tax rate that increases with income. In addition, the highest rates (at around 2%) are found in the South in the three regions (Molise, Campania and Calabria) subject to repayment plans for the deficit due to health care spending in 2012, while the lowest rates (at an average tax rate of around 1%) are observed in some northern regions where taxpayers in the lowest income bracket are either exempt or subjected to a reduced rate (Alto Adige and Friuli Venezia Giulia). The highest average tax rates are found in regions with higher incomes than the rest of the country and a progressive tax rate, as is the case of some northern regions (Emilia Romagna, Piedmont and Liguria) and one central region (Lazio).

Most of the regions with a prevailing single average rate settle at values around 1.23% and 1.73%, which are the basic statutory rates in 2012 established by the central government.

Figure 13 a - Additional regional tax by income tax bracket and geographical area *(average tax rate 2012[(a)])*

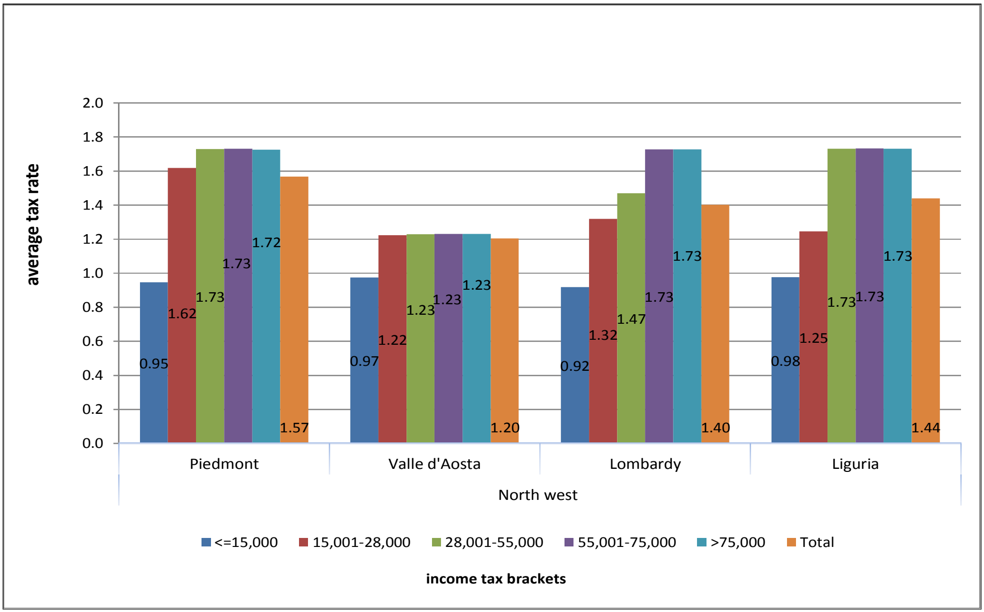


Figure 13 b - Additional regional tax by income tax bracket and geographical area *(average tax rate 2012[(a)])*

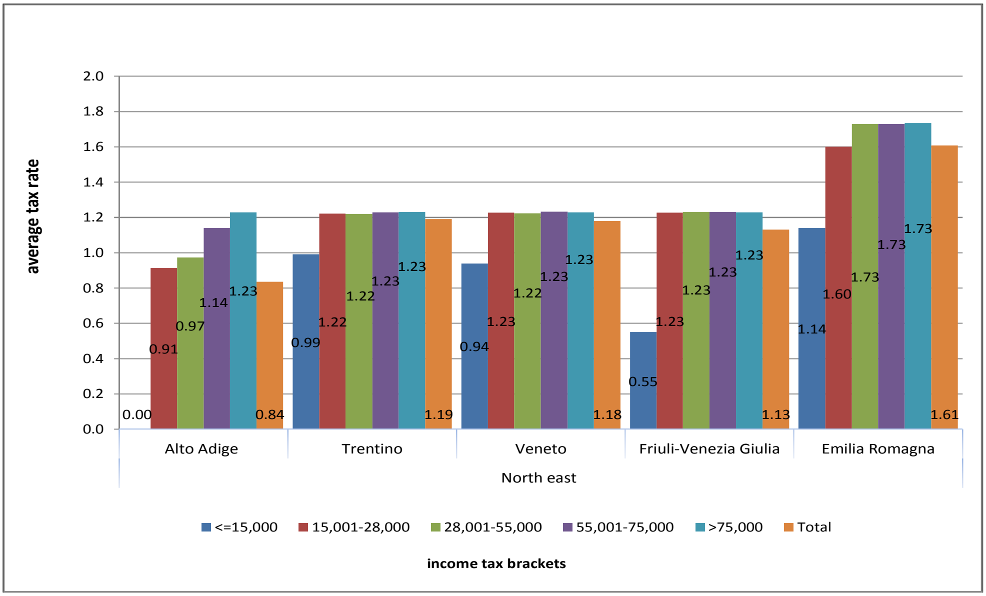


Figure 13 c - Additional regional tax by income tax bracket and geographical area *(average tax rate 2012[(a)])*

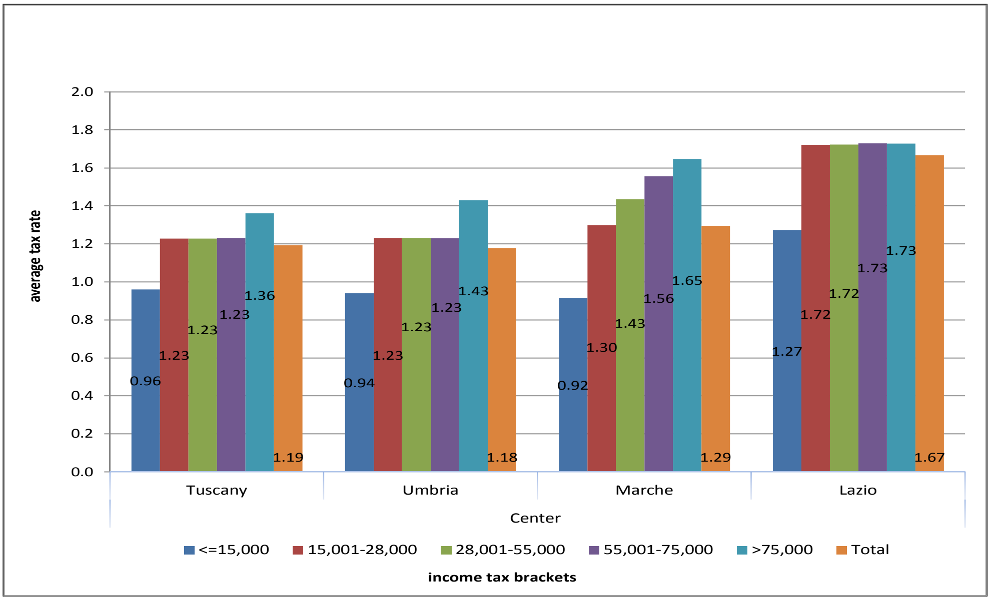
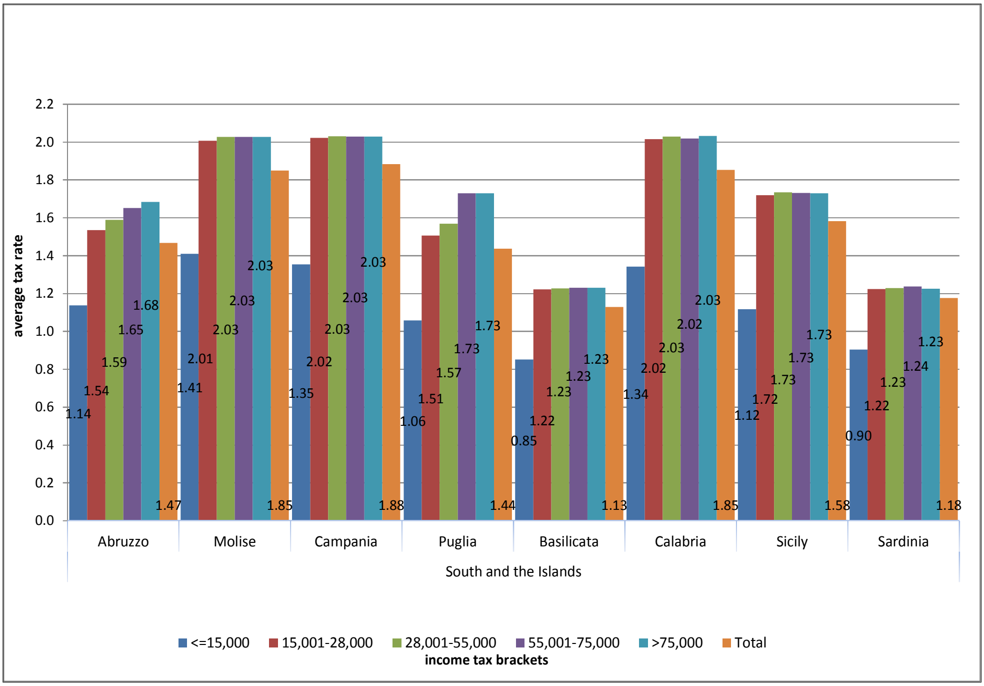


Figure 13 d - Additional regional tax by income tax bracket and geographical area *(average tax rate 2012[(a)])*



Source: Istat: Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012
(a) The average tax rate is calculated as the ratio between the revenue from the additional regional tax and the corresponding tax base.

Turning to the additional municipal tax, Figures 14 a and b show the average rate, that is the ratio between the revenue from the additional tax and the tax base according to the different personal income tax brackets (Irpef), by macro area in 2012. In all areas the additional municipal tax shows a proportional or mildly progressive structure, with a rate that increases marginally in all income brackets above the first (over 15,000 euros). There is a slightly greater increase in the rate in the last income bracket (over 75,000 euros) in the North-west, Center, South and the Islands. A mild case of regressive municipal tax is found in the South in the transition from the third to the fourth income bracket, in which the average tax rate is lower, although to a limited extent. Finally, it should be noted that the highest rates are found in the Center, probably due to the presence of the city of Rome[25], followed by the South, the Islands and the North-east.

[25] The city of Rome has many inhabitants almost half of the entire region and has also high additional tax rates.

Figure 14 a - Additional municipal tax by income tax bracket and geographical area *(average tax rate 2012[(a)])*

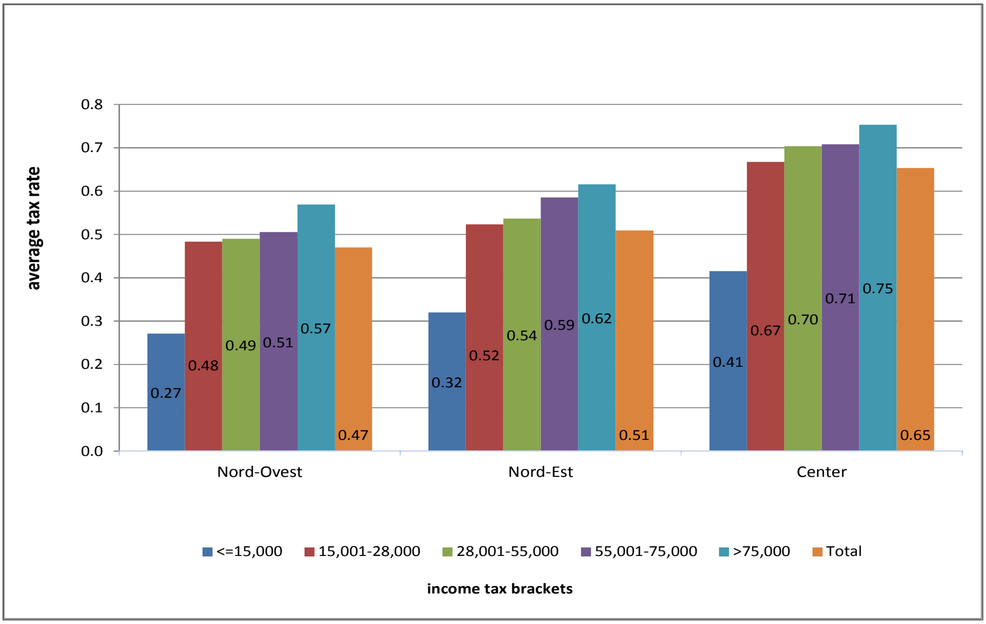
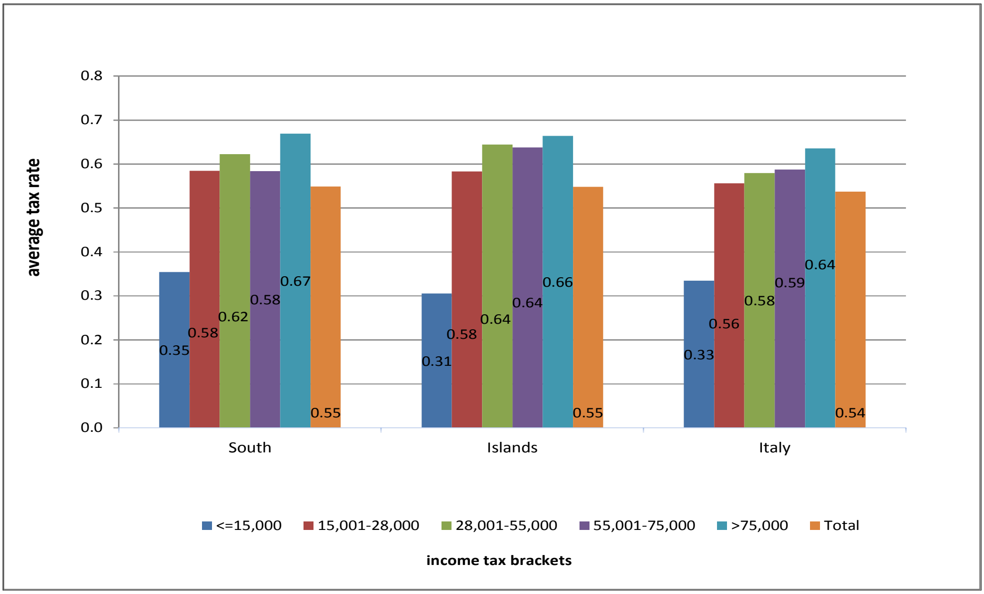


Figure 14 b - Additional municipal tax by income tax bracket and geographical area *(average tax rate 2012[(a)])*

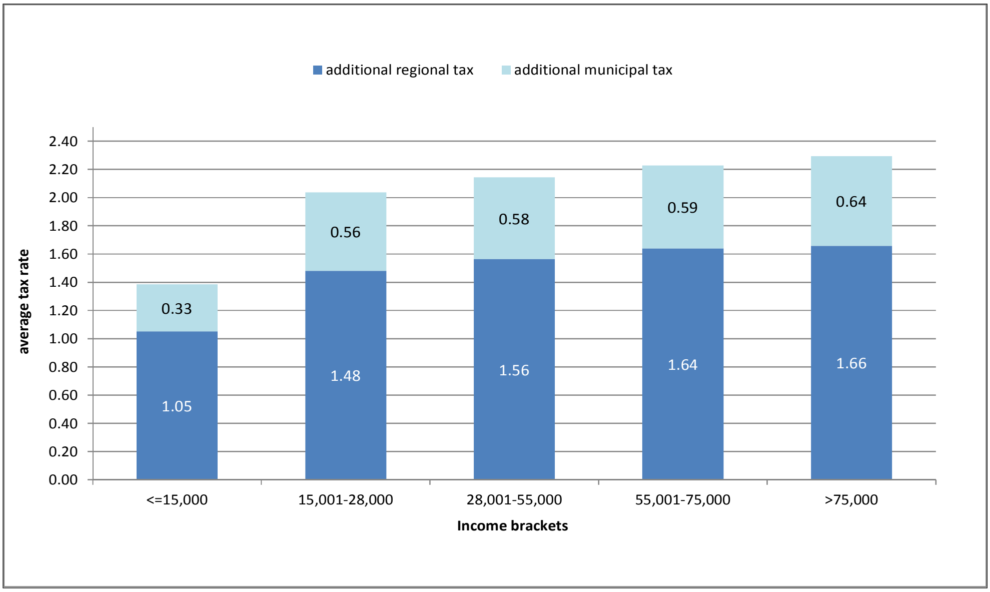


Source: Istat: Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012

(a) The average tax rate is calculated as the ratio between the revenue from the additional municipal tax and the corresponding tax base.

Finally, Figure 15 shows the average combined rate[26] of additional regional and municipal taxes in Italy, according to the different personal income tax brackets (Irpef) in 2012. The two taxes are moderately progressive: the most consistent rate increase, for both the regional and the municipal rate, occurs in the transition from the first to the second personal income tax bracket and, although to a lesser extent, in the transition from the second last to the last personal income tax bracket.

Figure 15 - Additional regional and municipal tax by income tax bracket – Italy - *(average tax rate 2012[(a)])*



Source: Istat: Microsimulation model of household taxes and benefits (FaMiMod): 2009 incomes updated to 2012
(a) The average tax rate is calculated as the ratio between the revenue from the additional regional and municipal tax and the corresponding tax base.

The result is a tax levy system that is complex and heterogeneous. The system has potentially distortionary effects due to the lack of coordination between the redistributive goals pursued at the central level and those introduced at the local level, as well as local differences between the various regions and municipalities, to the effect that taxpayers with the same income can be treated very differently depending on the region or the municipality in which they reside. Moreover, this system also increases the overall tax burden.

[26] The average tax rate is calculated as the ratio between the revenue respectively of the additional regional and of the additional municipal taxes and the corresponding tax base.

6. Family allowances

6.1 The Italian system

The current system of Family Allowances (*"Assegno per il nucleo familiare"*, hereafter ASF) can be viewed as the result of a troubled history, in which subsequent laws are stratified and overlap in time, without merging in a fully organic and coherent design.

The first economic support to family burden in Italy dates back to the mid-thirties[27]. Since then, a long sequence of interventions and micro-interventions have occurred over the decades: the 1955, 1988 and 2007 reforms attempted to rationalize and standardize this matter, not always fully achieving this goal. Just to sketch the long-term trends of this regulatory process, we can highlight some relevant aspects:

- the shift from a social security measure, categorical (i.e. related to the employment status of the recipient) and characterized by high funding rates, to a social assistance tool designed to support family income, financed by general taxation: recipients have been progressively extended – the measure remaining still categorical – including retirees (former employees), those receiving unemployment benefits and more recently (with 1998 Budget Law, no. 449/1997) the contract workers (termed Co.co.co or Co.co.pro,)
- the gradual merging of benefit structures, originally differentiated according to working sector[28];
- the introduction of selection rules (since 1988), varying allowance amounts on the basis of family income and excluding households beyond a maximum threshold[29];
- the introduction of benefit differentiation according to the number of family members, particularly children (making it more generous for larger families), and depending on the presence of disabled, single parents or orphans.

The 2007 Budget Law (no. 296/2006) outlined the current system: the most significant change, compared to before, is represented by the introduction of a "quasi-linear" decreasing rule in the amount of the allowances, in order to soften significantly the previous strong "poverty trap" effects (see Figure 16)[30]. The recipients were defined, according to Law no. 153/1988 (no amendments occurred subsequently on this respect), as families fulfilling certain requirements. These are briefly summarized below.

---

27 We shortly report some focal point of the story, a fully discussion falling outside the scope of the present contribution. More insights can be found in Ricci (2008).

28 Allowances were different depending on whether the worker was an employee in industry, commerce, etc.; civil servants benefited of the so-called "family addition", differentiated on the basis of the size of the municipality. The unification of all treatments occurred with the 1988 reform (Law no. 153/1988) that defined the "*Assegno per il Nucleo Familiare*". At present, only allowances for farmers, tenant farmers and sharecroppers, together with pensioners formerly self-employed, are diversified.

29 Actually, the first measures differentiating benefits with family income and establishing maximum threshold for access to the benefit refer to some years before (Law Decrees no. 17/1983 and no. 70/1984, Law no. 41/1986). However, the 1988 reform organically introduced ASF decreasing with income: mainly for this reason, the number of individuals living in beneficiary households showed a marked decrease during the eighties (from 24 to 14 million persons).

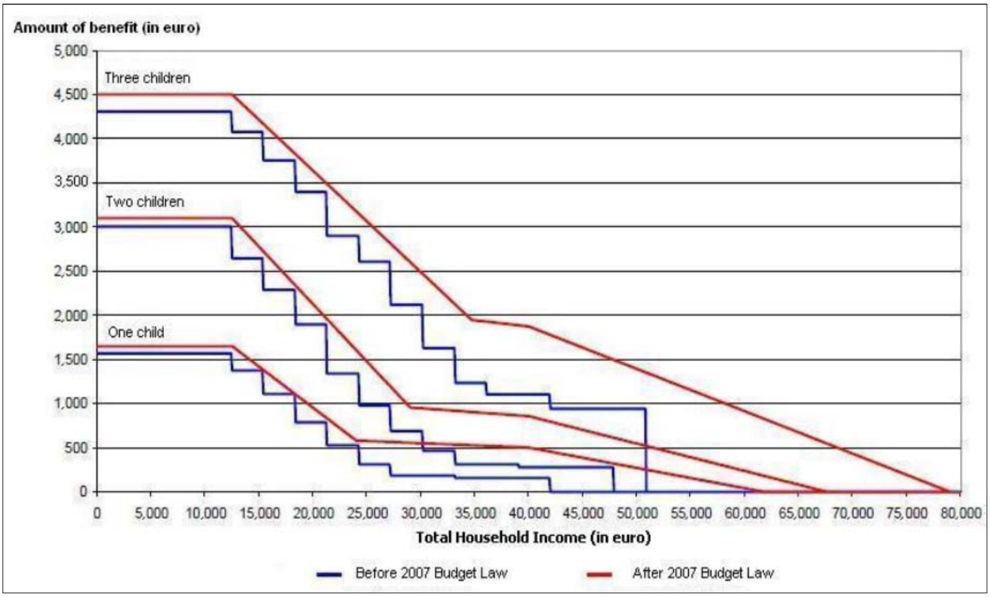
30 As results from Figure 16, taken from Ricci (2008), in 2006 the annual benefit for a three children family was equal to 942 euros for incomes up to 49,968 euros and became zero beyond this level, while for five children the benefit below the maximum income threshold (55,776 euros) even reached 2,268 euros.

Demographic requirements: the family unit is composed of the applicant, his/her spouse, children of minor age and, if disabled, even older (provided that they are not married). If there are at least four children under 26 years, children aged 18 or more but less than 21 are also included, as long as they are students or in training. Finally, brothers, sisters and grandchildren of the applicant are included, as are minors or disabled adults if they have lost both parents, are unmarried and not entitled to survivor pension.

Job requirements: employees (including part-time), unemployed receiving unemployment benefits, retired former employees, contract workers, domestic workers and farm workers are eligible for the benefit. The amount is paid monthly and is immediately cut off when any requirement is lost.

Income requirement: the sum of all incomes subject to the income tax of the eligible family members is considered, including (if above 1,032.91 euros) tax-free incomes or those subjected to substitute tax[31]. However, at least 70% of household income must result from wages (or incomes from contract work), unemployment benefits and pensions. Yearly income of the family in year *t - 1* determines the amount of the benefit for the second half of year *t* and the first half of *t + 1*.

Figure 16 - Family allowances before and after 2007 Budget Law *(Household with both parents and no disabled)*



Source: Ricci (2008)

31 Among the first, social pensions/allowances and disability pensions are to be included, whereas allowances for the attendance of disabled persons, war pensions and family allowances themselves are excluded. As for incomes subjected to substitute tax, the typical example is interest on financial assets.

6.2 Modelling allowances

Despite the considerable rationalization carried out by 2007 Budget Law, ASF scheme is still based on a complex mechanism of income thresholds, benefit values and decreasing paths, all depending on the number of eligible family members making up the family unit and some other household characteristics. One can briefly summarize the model using the following expressions and parameters:

- a first threshold s_ass_0 variable by family size and type, identifying the income level below which the benefit is maximum;
- a final threshold s_ass_{NO} variable by family size and type, identifying the income level above which the benefit is null;
- intermediate thresholds, also variable by family size and type, identifying the income level at which the decreasing rule (i.e., the speed) changes;
- an additional allowance for single-parent households with 3 or more children, also defined according to thresholds and decreasing rules varying by family size and type.
- In short, allowances can be expressed by[32]

$$ASF = ASF_0^F \, , \quad Y < s_{ass}{}_0^F$$

$$ASF = ASF_0^F - \left[\sum_{i=1}^{K-1} p_i^F \frac{\left(s_{ass}{}_i^F - s_{ass}{}_{(i-1)}^F\right)}{INT}\right] - p_K^F \frac{\left(Y - s_{ass}{}_{(K-1)}^F\right)}{INT}\, , \qquad s_{ass}{}_{(K-1)}^F \le Y < s_{ass}{}_K^F$$

$$ASF = 0 \, , \qquad Y \ge s_{ass}{}_{NO}^F$$

where *s_ass* and *p* are depicted in Tables 6 and 7, *K* may vary between 1 and 5 (depending on family type), *F* indicates parameter variability according to household characteristics (i.e., number of children, parents, presence of disabled persons) and the income bracket *INT* was fixed by 2007 Budget Law equal to 100 euros and yearly increases along with thresholds according to price index (currently it is equal to 110.27 euros).

[32] As already mentioned, the expression represents a proxy of the quasi-linear decreasing rule. The resulting values are coherent with official ones at the top of each income class; on the contrary, for the bottom values of each class the discrepancy is maximum, the size depending on the decreasing rule in that income class.

Table 6 – Family allowance Income thresholds, by family type *(2012 July - 2013 June)*

Family type	s_ass_0	s_ass_1	s_ass_2	s_ass_3	s_ass_4	s_ass_{NO}
both parents, at least one minor, no disabled, 3 persons	13,784.93	26,467.05	44,111.75			68,042.37
both parents, at least one minor, no disabled, 4 persons	13,784.93	31,981.03	44,111.75			74,548.84
both parents, at least one minor, no disabled, 5 persons	13,784.93	38,266.94	44,111.75			87,120.71
both parents, at least one minor, no disabled, 6 persons	13,784.93	23,489.51	39,810.85	49,625.72		89,767.40
both parents, at least one minor, no disabled, 7 persons	13,784.93	23,489.51	39,810.85	43,008.96	49,625.72	95,171.09
single parent, at least one minor, no disabled, 2 persons	13,784.93	26,467.05	44,111.75			68,042.37
single parent, at least one minor, no disabled, 3 persons	13,784.93	31,981.03	44,111.75			74,548.84
single parent, at least one minor, no disabled, 4 persons	13,784.93	38,266.94	44,111.75			87,120.71
single parent, at least one minor, no disabled, 5 persons	13,784.93	23,489.51	39,810.85	49,625.72		89,767.40
single parent, at least one minor, no disabled, 6 persons	13,784.93	23,489.51	39,810.85	43,008.96	49,625.72	95,171.09
single parent, at least one minor, no disabled, 4 persons - additional allow ance	15,990.51					28,812.60
single parent, at least one minor, no disabled, 5 persons - additional allow ance	15,990.51	58,448.07				91,722.11
single parent, at least one minor, no disabled, 6 persons - additional allow ance	23,489.51	61,756.44				105,633.00
both parents, at least one disabled, 3 persons	24,592.30					68,097.26
both parents, at least one disabled, 4 persons	24,592.30					74,564.37
both parents, at least one disabled, 5 persons	24,592.30					87,159.99
both parents, at least one disabled, 6 persons	30,878.22					89,806.39
both parents, at least one disabled, 7 persons	33,083.81					95,198.73
single parent, at least one disabled, 2 persons	24,592.30					68,097.26
single parent, at least one disabled, 3 persons	24,592.30					74,564.37
single parent, at least one disabled, 4 persons	26,797.89					87,225.08
single parent, at least one disabled, 5 persons	30,878.22					91,659.79
single parent, at least one disabled, 6 persons	33,083.81					105,546.95
single parent, at least one disabled, 7 persons	36,392.19					109,636.05

Source: INPS

Table 7 – Family allowance parameters, by family type *(2012 July - 2013 June)*

Family type	ASF_0	p_1	p_2	p_3	p_4	p_5
both parents, at least one minor, no disabled, 3 persons	1,650	9.3	0.5	2.3		
both parents, at least one minor, no disabled, 4 persons	3,100	13.0	0.9	3.1		
both parents, at least one minor, no disabled, 5 persons	4,500	11.5	1.4	4.8		
both parents, at least one minor, no disabled, 6 persons	6,000	5.0	10.5	19.6	6.2	
both parents, at least one minor, no disabled, 7 persons	7,500	7.5	11.2	1.6	25.0	8.8
single parent, at least one minor, no disabled, 2 persons	1,650	9.3	0.5	2.3		
single parent, at least one minor, no disabled, 3 persons	3,100	13.0	0.9	3.1		
single parent, at least one minor, no disabled, 4 persons	4,500	11.5	1.4	4.8		
single parent, at least one minor, no disabled, 5 persons	6,000	5.0	10.5	19.6	6.2	
single parent, at least one minor, no disabled, 6 persons	7,500	7.5	11.2	1.6	25.0	8.8
single parent, at least one minor, no disabled, 4 persons - additional allow ance	1,000	8.6				
single parent, at least one minor, no disabled, 5 persons - additional allow ance	1,000	1.5	1.4			
single parent, at least one minor, no disabled, 6 persons - additional allow ance	1,550	1.6	2.5			
both parents, at least one disabled, 3 persons	2,020	5.1				
both parents, at least one disabled, 4 persons	3,920	8.7				
both parents, at least one disabled, 5 persons	5,640	9.9				
both parents, at least one disabled, 6 persons	7,690	14.4				
both parents, at least one disabled, 7 persons	9,700	17.2				
single parent, at least one disabled, 2 persons	2,020	5.1				
single parent, at least one disabled, 3 persons	3,920	8.7				
single parent, at least one disabled, 4 persons	6,280	11.5				
single parent, at least one disabled, 5 persons	8,450	15.3				
single parent, at least one disabled, 6 persons	11,040	16.8				
single parent, at least one disabled, 7 persons	13,590	20.5				

Source: INPS

6.3 ASF: simulation issues

As is known, microsimulations generally need to be translated into working hypotheses, attempting to approximate current laws (or an alternative, hypothetical, scenario). In some cases, one can reach a good level of approximation of the "real world", whereas in others some puzzles remain unsolved. Let us here review the problems that arose and the solutions given – when feasible – in the simulation.

A first issue concerns the definition of the family that is relevant for ASF: families in the model were derived from EU-SILC households, recoding kinships when appropriate to identify nuclear families within the households; therefore, in some household there can be two or more families that can apply for ASF.

A second issue relates to disability: to be eligible as a ASF recipient, a disabled person "has absolute and permanent inability to engage in work, for physical or mental dearth". In the model, we count as disabled the pensioners and housewives who receive a disability pension (without labour income).

Using FaMiMod, the number of family allowances is estimated at 4.8 millions for

2012, with total expenditure equal to 5.2 billion euros (the resulting average allowance is 1,071 euros). Most benefits (96% of beneficiaries, 94% of expenditure) are paid to families with children and without disabled persons. Instead, recent National Accounts showed that family allowances in 2012 amounted to 6.4 billion euros[33]. The underestimation (over one billion euros) could be partly due, to the assumption that family income is entirely (i.e. honestly) reported when applying for ASF benefits, with the only exception of financial incomes.

Hence, an improvement in the treatment of tax avoidance is required fill the gap with National Accounts figures.

Further improvements may come from the simulation of family allowances for farmers, tenant farmers and sharecroppers as well as those directed to surviving spouses.

7. What is next?

An ambitious task of Istat is to jointly provide, for any given tax/policy, a representation of its redistributive effects and of the expected aggregate change in the public budget. The available data and the micro-simulation techniques, encompassed in FaMiMod, do not allow the simultaneous estimation of the redistributive effects and of total tax revenues. A second best *ad hoc* solution would be to setup an additional calibration of weights to compute the total amount of expenses and/or revenues. Such a calibration should correct the wide difference in the amount of under-reported income in administrative with respect to survey data. A more refined solution requires to find an appropriate mix of different analytical strategies:

- integration of FaMiMod with the macroeconomic forecasts provided by the Istat Me-mo model;
- in depth analysis of the whole information about the universe of taxpayers. On the one hand, it would be important to assess the statistical representativeness of the EU- SILC theoretical sample (*i.e.* including the non-respondents) when the reference population is the totality of Italian taxpayers. A first advantage would be the inclusion of the aggregate amounts from the tax files as constraints in the weighting procedure. On the other hand, the study of the differences between the incomes reported in the two sources of microdata would allow an improvement of the proxy measure of tax avoidance, which is now too simple;
- setup of a dynamic micromodel, including behavioural responses and demographic ageing, to obtain a better updating of both administrative and survey microdata from the base year to the desired date and to account for the expected individual reactions to policy changes.

[33] See dati.istat.it → National Accounts → Environmental and other satellite accounts → Social Protection Accounts. Actually, the 2012 value is slightly higher, amounting to 6.580 billion euros, since some items are delivered by INAIL (National Institute for Work Accidents Insurance). Since information on the total number of cheques paid is missing, we can consider data referring to year 2011 as a proxy (see 2011 General Report on Economic Situation in the Country): for that year, the number of beneficiaries in the private sector amounted to 3,961,000 (source: INPS), to which allowances provided to public workers must be added, which may be approximately estimated at about 400,000.

References

Baldini, M., E. Ciani and D. Pacifico (2011), *Mapp, a Tax Benefit Microsimulation Model for the Analysis of Public Policies in Italy,* University of Modena and Reggio Emilia, Department of Economics.

Bourguignon, F. and A. Spadaro (2006), Microsimulation as a tool for evaluating redistribution policies, *Journal of Economic Inequality*, 4, 1, 77-106.

Ceriani, L., C.V. Fiorio and C. Gigliarano (2013), The importance of choosing the data set for tax-benefit analysis, *The International Journal of Microsimulation*, 6(1), 86-121.

Creedy, J. (2003), Survey reweighting for tax microsimulation modelling, *Treasury Working Paper Series* 03/17, New Zealand Treasury.

Deville, J.-C. and C.-E. Särndal (1992), Calibration estimators in survey sampling, *Journal of the American Statistical Association*, 87, 418, 376-382.

Di Biase, R., M. Di Marco, F. Di Nicola and G. Proto (1995), ITAXMOD: A Microsimulation Model of the Italian Personal Income Tax and of Social Security Contributions, *Documenti di Lavoro ISPE*, n.16/95, Ispe, Roma.

Di Marco, M. (2007), Self-Employment Incomes in the Italian EU-SILC: Measurement and International Comparability, in *Comparative EU statistics on Income and Living Conditions: Issues and Challenges*, Eurostat, Luxembourg, 145-156.

Gomulka, J. (1992), Grossing-up revisited, in R. Hancock and H. Sutherland (eds), *Microsimulation Models for Public Policy Analysis: New Frontiers*, Suntory-Toyota International Centre for Economics and Related Disciplines, London School of Economics and Political Science, London, 121-132.

Immervoll, H., K. Lindström, E. Mustonen, M. Riihelä and H. Viitamäki (2005), Static data ageing techniques: accounting for population changes in tax-benefit microsimulation models, *EUROMOD Working Paper*, EM7/05.

Maitino, M. L., L. Ravagli and N. Sciclone (2013), Il modello di microsimulazione dell'Irpet MicroReg: aggiornamenti ed estensioni, *Studi e approfondimenti IRPET*, June.

OECD (2012), *Economic Outlook*, Volume No. 92, OECD Publishing, Paris.

OECD (2013), *Taxing Wages 2013,* OECD Publishing, Paris.

Proto, G. (1999), Il modello di microsimulazione MASTRICT: struttura e risultati, *Rivista di Statistica Ufficiale,* n. 3, Istat, Roma.

Ricci, L. (2008), Gli assegni per il nucleo familiare: una storia parallela, in C. De Vincenti and R. Paladini (eds.), L'imposta sui redditi delle persone fisiche e il sostegno alle famiglie, *Tributi*, supplemento n. 1.

Sutherland, H. and F. Figari (2013), EUROMOD: the European Union tax-benefit microsimulation model, *The International Journal of Microsimulation,* 6(1), 4-26.

Redistributive effects of changes in indirect taxation[1]

Alessandro Brunetti[2], *Maria Grazia Calza*[3]

Abstract

The aim of the paper is to outline the methodological solution adopted in modelling the effects of changes in indirect taxation on household expenditure in Italy. More specifically, an analysis is carried out on the results of the application of a preliminary version of the model with a focus on the evaluation of the effect of the tax reforms introduced in 2011 to support the consolidation process of public accounts. Finally, the effects on household expenditure deriving from the increase of the standard value added tax (VAT) rate (from 20% to 21%) and of fuel excise duties are discussed.

Keywords: Value added tax, fuel excise duties, household expenditure.

JEL Code: H20, H22, H23

1. Introduction

In the last decades, microsimulation models have been extensively used in evaluating the redistributive effects of public policies, due to the gain in terms of accuracy of the analysis offered by models which consider explicitly heterogeneity of the agents.

This paper presents the microsimulation model developed in Istat with the aim of analysing effects of changes in indirect taxation on the households expenditure in Italy. According to the taxonomy proposed in Bourguignon and Spadaro (2006) the model is an *arithmetical* one, since it ignores any behavioural reaction by households to a change in indirect taxation.

The paper is organized as follows. Microdata used to build the model and the underlying assumptions are outlined in paragraph 2, where the equations that specify the model are also introduced. In paragraph 3, the results of the application of a preliminary version of the model for the evaluation of the effect of the tax reforms introduced in 2011 are discussed. Moreover, the analysis investigates the change in the composition of households expenditure over the period 1997-2011, by VAT rates and quintile expenditure groups. Paragraph 4 concludes by sketching out the directions for the future development of the model.

[1] The authors are grateful to Mariangela Zoli and to the anonymous refereers for useful comments to a preliminary draft of the paper. Although the paper reflects the joint work of the authors, Alessandro Brunetti is responsible for paragraphs 1, 2 e 3 while Maria Grazia Calza for paragraphs 4 and 5. The view expressed in this paper are solely those of the authors and do not involve any responsibility of Istat.

[2] Researcher (Istat), e-mail: albrunet@istat.it.

[3] Senior Researcher (Istat), e-mail: calza@istat.it.

2. Building the model: data and hypothesis

To derive microdata input for simulation we use the dataset of the Household Budget Survey (HBS) by Istat. The HBS, in fact, provides information on household expenditures for consumption (good and services), including own final consumption data, goods and services provided directly to workers by employers, imputed rental costs from owner-occupied houses and free rental cost houses. Since 1997, when the survey design was completely renewed, the Italian HBS has been providing time series data on yearly basis[4].

Specifically, in 2011 the survey involved 28,000 households having residence in Italy representative of a total of 25,165,002 households with an average monthly expenditure equal to 2,487.91 euros. The collected information has been organized in a data set containing, at the individual level, the main socio-demographic characteristics of households and their consumption expenditures referred to 277 aggregates of products.

The aim of estimating the effects of changes in indirect taxation on households budget, requires, as a preliminary step, to link each single expenditure to the appropriate VAT rate[5] and, for fuels, to the excise duty.

The link between HBS aggregate expenditures and fiscal parameters is carried out by using the information collected by Consumer price index survey for the calculation of the Harmonized Index of Consumer Price at Constant Tax. The HICP-CT is a satellite index of the headline inflation indicator calculated by ISTAT according to the European Union regulations in force and which is used to estimate the development of consumer prices on a comparable basis at EU level[6]. Precisely, the HICP-CT is defined as an index where tax rates are kept constant so that, in the event of a tax rate change, the difference between the current HICP-CT and HICP would measure the effect of the tax rate change on final prices, under the hypothesis that any rise (decline) in tax rates is completely and suddenly passed on as a price increase (decrease). For the compilation of the Italian HICP-CT, the tax rates levied on more than 600 products included in the basket of the HICP are monthly monitored, together with the excise duties on fuels and tobacco products.

To associate each HBS expenditure aggregate with the appropriate tax rate (or tax rates), all the items in the HICP basket are grouped according to the HBS classification structure. As a result, the following two possible cases occur:

1) One or more products in the HICP basket, charged with the same VAT rate, are associated to a single HBS aggregate;
2) More than one product in the HICP basket, charged with different VAT rates, are associated to a single HBS aggregate;

Generally, the products classified in a single HBS aggregate are all subjected to the same (standard or reduced) tax rate, they are all exempted or are out of the scope of VAT

---

4 From 2005 the survey has included a new variable that allows to classify households as poor or not poor according to the definition of absolute poverty.

5 In Italy, the Value added tax is split in three rates. In 2011 the tax rates were 4%; 10% and 20%. In September of the same year, the standard rate was increased to 21% . More recently (October 2013) another increase pushed the standard rate to 22%. However, it should be noted that certain goods and services are exempt from VAT (for example, postal services, medical care, insurance). For these cases, the VAT tax rate is considered to be equal to 0.

6 More details on HICP-CT and more generally on the Consumer price indices survey can be found in (Istat 2013a).

(as for imputed rents). However, in a number of cases, the link is “one to many”: it is, in fact, possible that different VAT rates are levied on the prices of different products included in the same HBS aggregate expenditure. This is the case, for example, of fresh and UHT milk, whose final prices include respectively 4% and 10% VAT rates or, to make another example, the case of TV and Pay-TV subscriptions which include in turn 4% and 21% VAT rates (22%, after 1st October 2013).

While in the first case it is possible to link unambiguously the appropriate tax rate to HBS aggregates, in the second one the relative weight of different tax rates linked to the same aggregate have to be estimated[7]. Or equivalently, a weighted average tax rate has to be calculated. It is easy to show that the weight of each tax rates should be proportional to the ratio of the corresponding tax base on the total. Formally, let x_i be the expenditure for the *i*-th aggregate of products. For the sake of simplicity (but the argument can be extended to the general case) let us assume that only two products are included in aggregate *i* whose prices are charged with two different VAT rates. That is:

$$x_i = x_{i,1} + x_{i,2} = \tilde{x}_{i,1} \cdot (1 + \alpha_1) + \tilde{x}_{i,2} \cdot (1 + \alpha_2)$$

where the tilde is used to denote the tax base and α_1 and α_2 are the two VAT rates. In this case, it is possible to express x_i as follows:

$$x_i = \tilde{x}_i \cdot (1 + \bar{\alpha}_i)$$

where:

$$\tilde{x}_i = \tilde{x}_{i,1} + \tilde{x}_{i,2}$$

and

$$\bar{\alpha}_i = \frac{\tilde{x}_{i,1}}{\tilde{x}_i} \cdot \alpha_1 + \frac{\tilde{x}_{i,2}}{\tilde{x}_i} \cdot \alpha_2 = w_{i,1} \cdot \alpha_1 + w_{i,2} \cdot \alpha_2$$

According to the last expression, in order to calculate the weighting coefficients for VAT rates, the corresponding tax bases have to be determined. To this aim, we exploit the consumption expenditure estimates used within the HICP framework for the computation of the weights assigned to the products in the basket of the index. It should be noted that these consumption estimates refer to the whole population. Therefore, their use in the microsimulation model introduces the implicit assumption that, within the HBS expenditure aggregates, the ratio of tax bases is constant across households.

A second issue addressed by the model concerns the monetary effects of a change of the excises on fuels. Since these are per unit taxes, to estimate such effects, households expenditure for fuels need to be decomposed into price and quantity components. To this

[7] Moreover, since a limited number of HBS aggregate expenditures are not included in the domain of the HICP (such as life insurance or major repairs connected with dwelling) the corresponding VAT rates are defined using other sources of information.

aim, data concerning average prices of unleaded petrol, Diesel oil and liquefied petroleum gas (GPL), released by the Italian association of petroleum companies (Unione Petrolifera) are used.

3. Modelling the effect of indirect tax changes

This section introduces the equations used to estimate the effect of the indirect taxation change on households' expenditures. The literature on microsimulation models distinguishes two different approaches according to (on the basis of) their underlying hypotheses about the behaviour of economic agents. In *behavioural* models, agents are assumed to modify their decisions when the variables defining their economic environment change. In this approach, households are supposed to adjust their consumption pattern in response to a change in indirect taxation. At the opposite, in the *arithmetical* approach, reactions by households are ruled out of the model[8]. This is also the approach adopted by the module of Istat microsimulation model dealing with indirect taxation. More precisely, the equations of the model are based on two main hypothesis, which may be both considered a consequence of the assumption of rigid consumers demand (zero elasticity of substitution):

1) the quantities purchased by households remain constant when tax rates are changed;
2) any change in indirect taxation is completely passed on final prices.

In what follows, we will address the VAT case first and then present the formula for the evaluation of the impact of changes in the excises on fuels.

Let α_j^0 and α_j^1 be the VAT rates in force in period 0 and 1 respectively, where $j = 1, \ldots, 3$ refers to the VAT rates class[9].

Since the consumption decisions of the households are supposed to be unaffected by fiscal policies, the expenditure change for the *i*-th aggregate between period 0 and 1 is then estimated as follows[10]:

$$\Delta x_i = x_i^1 - x_i^0 = \tilde{x}_i \cdot (\bar{\alpha}_i^1 - \bar{\alpha}_i^0)$$

or equivalently,

$$\Delta x_i = x_i^0 \cdot \sum_j w_{i,j} \cdot \frac{\alpha_j^1 - \alpha_j^0}{(1 + \bar{\alpha}_i^0)}$$

where, as before:

[8] The conditions under which the arithmetical approach can be theoretically justified are discussed in Bourguignon and Spadaro (2006)

[9] That is, super-reduced rate (j=1), reduced rate (j=2), standard rate (j=3).

[10] Notably, by ignoring behavioral responses by households, the weights used to calculate the average VAT rate do not depend on the value of α_j.

$$\bar{\alpha}_i^t = \sum_f w_{i,f}\, \alpha_f^t \qquad t = 1, 2$$

and

$$w_{i,f} = \frac{x_{i,f}}{x_i}$$

It is important to note that, in this framework, the increase (decrease) of households expenditure corresponds to the rise (decline) of tax revenues:

$$\Delta x_i = \sum_f w_{i,f} \frac{x_i^0}{(1+\bar{\alpha}_i^0)} \cdot \alpha_f^1 - \sum_f w_{i,f} \frac{x_i^0}{(1+\bar{\alpha}_i^0)} \cdot \alpha_f^0$$

Concerning the effect of excises on fuels, let x_f^0 denote the expenditure for fuel f in period 0 and p_f^0 the corresponding gross price. p_f^0 is given by:

$$p_f^0 = \left(\tilde{p}_f^0 + e_f^0\right) \cdot (1 + \alpha_f^0)$$

where $\tilde{p}_f^0$ is the net price, e_f^0 and α_f^0 are respectively the excise and VAT rate.
Accordingly, the expenditure change between period 0 and 1 can be expressed as follows:

$$\Delta x_f = \left(\frac{p_f^1}{p_f^0} - 1\right) \cdot x_f^0$$

with p_f^1 given by:

$$p_f^1 = \left(\frac{p_f^0}{1+\alpha_f^0} + \Delta e_f\right) \cdot (1 + \alpha_f^1)$$

In the next section, we discuss the result of the simulation carried out on a preliminary version of the model to estimate the effect of the increase of the standard value added tax rate and the changes of the fuel excise duties that have been recently introduced in Italy[11].

4. Microsimulation results

Microsimulation analysis has been focused on the quantitative assessment of the impacts on household expenditure of some indirect tax reforms that have come into force

[11] The architecture of the model has been developed in Stata environment by Corrado Pollastri and Alessandro Brunetti.

during 2011. In detail, analysis has focused on the increase of the standard VAT rate that has come into force on September 2011 modifying the standard rate from 20% to 21%, and on the increase in some excise duties on motor fuel as specified in Table 1.

Table 1 – Main Indirect Tax Reforms. 2011-2012

Indirect tax	reform
VAT	standard rate increase from 20% to 21%
Excise duty on petrol	excise rate increase from EUR 564.00 to 728.40 on 1000 liters
Excise duty on heavy fuel oil	excise rate increase from EUR 423.00 to 617.40 on 1000 liters
Excise duty on LPG	excise rate increase from EUR 125.27 to 147.27 on 1000 liters

The impact on household expenditure has been estimated ranking households according to their equivalised expenditure[12]. Moreover, households have been divided into five groups of equal size. The bottom quintile group represents households with the lowest amount of expenditure while the top quintile represents households with the highest amount of expenditure. As specified above, estimates are made under the assumption of constant pre-tax prices and constant quantity of goods purchased. Thus, the tax increase is completely passed forward to consumers[13].

In detail, Figure 1 shows the total impact of the tax rise due to the various reforms. The bottom quintile group of households pays less in tax rise as a percentage of their total expenditure than the top quintile group (respectively 0.78% and 0.86%). As to the other three quintile groups, under the stated assumption of constant quantities, the increases in expenditure (from 0.88 to 0.92) are higher than the increase bearing on the top quintile group. However, if we disaggregate the total expenditure increase in the two components due to the VAT rise and the excise duty rise, results show a different distribution among quintile groups. In detail, looking at the increase of the excise duties, the tax incidence on the household expenditure is higher for the bottom quintile group than the top group (respectively 0.53 and 0.48%). As to the other groups they all bear an equal increase in total expenditure (0.60%) which is as well higher than the increase of the top group. Thus the excise duty increase is slightly regressive as it hits the poorest harder.

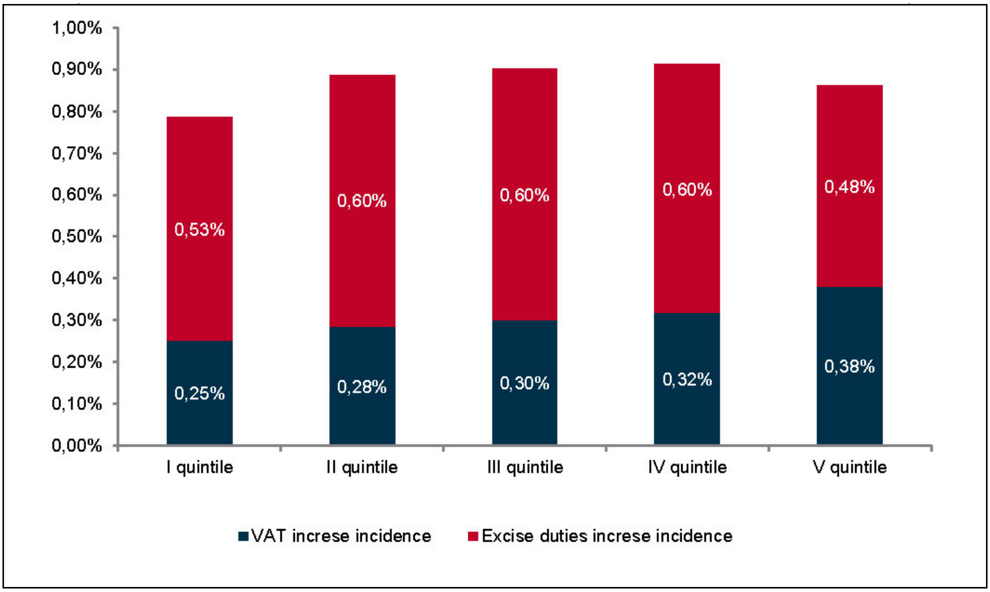
With regard to the incidence of the VAT standard rate, results show that incidence increases for higher level of total expenditure. The percentage increases in total expenditure range from 0.25 of the bottom group to 0.38 of the top group. Thus estimates indicate a progressive path for the five quintile groups, since the VAT increase has affected goods that are bought in a lower proportion by poorest households compared to the richer groups. In fact, the VAT increase does not affect the reduced rates which are applied to goods (such as

[12] Equalisation is used in order to adjust expenditure according to differences in household size. The equivalence scale used is the Carbonaro scale (1985).

[13] Recent studies on the price impact of tax reforms on consumption (Carbonnier, 2007) show that the tax increase could be partially absorbed by producers. Results of our estimates about the impact on different groups of household would be confirmed if the producer behaviour were equal among household groups.

food, housing, health and education) consumed in a higher proportion by poorest households (see Table 2).

Figure 1 – Incidence on total expenditure of increases in the standard VAT rate from 20 to 21 per cent and excise duties: by equivalised household expenditure groups in 2011



Source: **Household Budget Survey, estimates - Istat**

On the contrary, as shown in Table 2, the increase of the excise duties on fuels affects goods that are bought in a higher proportion by the bottom quintile group with respect to the top quintile. This causes the higher incidence on total expenditure for the bottom group, showing a regressive impact of this part of the tax reform.

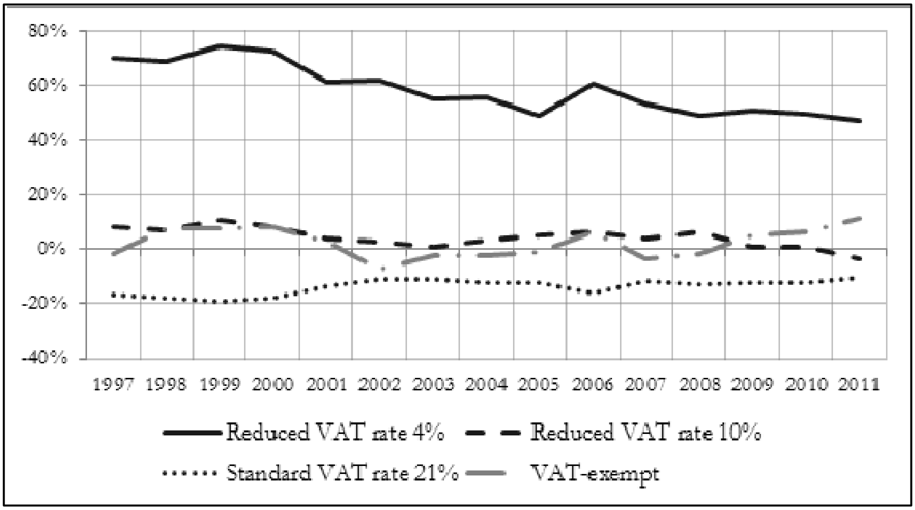
Nevertheless, increases in VAT rates can also be regressive if we take into account the changing spending patterns of poorest households during the last years. To this effect, analysis has investigated the changing proportion of household expenditure on standard VAT rate items and reduced VAT items by household groups during the period from 1997 to 2011. Specifically, it has been calculated the percentage deviation between the proportion of household expenditure on different VAT rate of the bottom quintile group of households with respect to the top group.

With reference to standard VAT rate items, during the considered period Figure 2 shows a converging path among the two household groups, that is the value of the percentage deviation decreases. Analysis shows also a decrease in the percentage deviation related to the reduced VAT rate (4%) items. Actually, in the past poorest households were used to allocate a high proportion of their expenditure on these items, but nowadays this proportion is lower. As to those VAT-exempt items and as well as for 10 per cent reduced VAT rate items, during the examined period the percentage deviation value between the bottom and the top quintile groups is always quite small.

Table 2 – Proportion of household expenditure on different VAT rate items and on total excise duty items: by equalised household expenditure groups in 2011

Expenditure quintile groups	VAT 4%	VAT 10%	VAT 21%	VAT-exempt	Imputed house rental costs	Total	share of fuels on total expenditure
I quintile	12.3	25.9	26.6	13.9	21.3	100.0	5.4
II quintile	10.8	24.4	29.9	11.2	23.7	100.0	5.6
III quintile	9.8	24.5	31.3	10.2	24.3	100.0	5.4
IV quintile	9.0	24.6	32.9	10.9	22.7	100.0	5.1
V quintile	6.7	25.2	39.4	10.8	17.9	100.0	3.8
Total households	8.8	24.9	34.0	11.0	21.2	100.0	4.8

Figure 2 – Percentage deviations between the bottom and the top quintile groups' expenditure proportions on different VAT taxed items – years 1997-2011



In conclusion, analysis has shown that due to these changing spending patterns the Italian VAT rate structure seems to be less able to protect poorest households by assuring a progressive or neutral impact of these tax reforms.

5. Conclusions and further research

Results from this microsimulation study highlight the potentials of this analytical tool drawing also the attention on future research needs. In particular, analysis is built on two main assumptions: the first assumes that the quantity of a good purchased remains constant after a change in taxes on expenditure and the second assumes that the pre-tax price remains the same, that is the full effect of a tax increase is passed on to the consumer. Removing these assumptions, research could investigate how tax changes affect price in different markets referring also to studies on purchase price dynamics and developing behavioral models for determining consumer reactions to changes in taxes on expenditure and thus in retail prices. In such a way this tool of analysis could be used to assess the impacts of public policies in specific markets or policies to guide consumer expenditure.

Moreover, the measurement of tax returns from changes in sales taxes requires the development of behavioral models for the economic agents that take also into account tax evasion and are linked with macro data from the national accounts.

Finally, changes in indirect taxes do not directly affect household incomes but they affect the amount of consumption and consequently they can affect the general level of economic welfare of the household. However, a better assessment of the distributional impacts of tax reforms requires the measurement of the impacts both on household expenditure and income. To this effect it would be important to integrate in a comprehensive database detailed survey information on household expenditure and on household disposable income.

References

Bernardi, L. (2013), Recent findings regarding the shift from direct to indirect taxation within the EA-17, *Siep Working paper*. 663 (June 2013).

Bourguignon, F. and A. Spadaro (2006), Microsimulation as a Tool for Evaluating Redistribution Policies, *ECINEQ Working paper*. 20.

Carbonaro, G. (1985), Nota sulle scale di equivalenza, in: *Commissione d'indagine sulla povertà e sull'emarginazione*, La povertà in Italia, Rome.

Carbonnier, C. (2007), Who pays sales taxes? Evidence from French VAT reform, 1987-1999, *Journal of Public Economics*. 91.

Caspersen, E. and G. Metcalf (1994), Is a Value added tax regressive? Annual versus lifetime incidence measures, *National Tax Journal*, vol. 47, no. 4 (pp. 731-46).

Istat (2013a), *Indici dei prezzi al consumo. Aspetti generali e metodologia di rilevazione*, Istat, Rome [http://www.istat.it/it/files/2013/04/Indice-dei-prezzi-al-consumo.pdf].

Istat (2013b), *Rapporto Annuale 2013. La situazione del Paese*, Istat, Rome [http://www.istat.it/it/files/2013/05/Rapporto_annuale_2013.pdf].

ONS (2011), Poorest households spending more on VATable items than in 1986, *Working paper Office for National Statistics*, UK.

Multi-source data collection strategy and microsimulation techniques for the Italian EU-SILC[1]

Paolo Consolini[2], Gabriella Donatiello[3]

Abstract

This chapter presents the multi-source data collection strategy that has been developed at the Italian National Institute of Statistics since 2004 for the EU-SILC project with a focus on the integration methodology that has been implemented to build net and gross income target variables. The first part of the paper describes the imputation and correction processes carried out by Istat to obtain the final income variables. The second part of the study explains the complex and innovative methodology devised to setup and use a microsimulation model when multiple integrated data sources are available, a task that goes far beyond the traditional "gross to net" (or "net to gross") conversion of survey incomes. The results show that combining microsimulations with integrated survey and administrative data definitely enhances data quality.

Keywords: Administrative Data, Survey Data, Data Integration, Microsimulation, Income, Multi-mode data collection, Record linkage.
C810 - Methodology for Collecting, Estimating, and Organizing Microeconomic Data

1. Introduction

The Italian SILC survey (EU-SILC) is based on the "Computer Assisted Personal Interviewing" method of collecting data and uses administrative microdata in order to reduce measurement errors. Many researchers, including statisticians, psychologists, sociologists and economists, share common concerns about the weakness of the measurement process in the survey method. As is well known, errors can be due to any of the many factors influencing the measurement process: the questionnaire, the respondent and the interviewer, as well as the data collection method. The structure and the wording of the questions affect the interpretation by the respondent. Even when the interviewee fully understands a question, he could still have memory problems in giving a reliable answer.

Understanding and memory problems often lead to measurement errors: omissions and recall errors being typical examples. In order to limit the possible bias on the income

[1] Paragraphs 2, 2.1, 2.2 have been drafted by Paolo Consolini; paragraphs 3, 3.1, 3.2 by Gabriella Donatiello and paragraphs 1 and 4 by both authors. The opinions are those of the authors and do not imply any responsibility for the Istat (Italian National Institute of Statistics).

[2] Senior Researcher (Istat), e-mail: consolin@istat.it
[3] Senior Researcher (Istat), e-mail: donatiel@istat.it

reported in the questionnaire by the interviewees and to improve the general data quality of the survey, a project of multi-source data collection has started at Istat since 2004. The integration technique used to combine survey and administrative microdata to produce the EU-SILC income target variables, can be viewed as a flow process, starting from the analysis of the income definitions adopted by the different data-sources, developing through the choice of the best matching key and the more effective record linkage methodology, followed by a consistent, problem-solving, approach for the harmonizing of units and variables, the handling of inconsistencies and of under/over coverage of the integrated data sources (survey, administrative, imputed, microsimulated) to end up with the reconciliation of values reported in the different sources with the final set of income target variables of the EU-SILC project. For the first Italian edition of the project (2004), the process involved only two ‘problematic’ income components: self-employment income and pensions. From the second edition (2005) onward it includes employment incomes, too.

In the following all the steps of the integration process will be analysed, focussing on the solutions adopted to handle the problems arising from the integration of different data sources (harmonization of units and definitions, incoherencies of income sources, reconciliation of inconsistent income amounts). At the same time, the impact that the data integration and editing process has on the final values of the income components will be provided and discussed. To sum up, the administrative data are used to support the editing and imputation processes and to ease the construction of gross incomes with microsimulation techniques.

According to EU Regulation, in Italy the estimation of gross income statistics became mandatory starting from survey year 2007. A microsimulation model which estimates taxes and social insurance contributions for the income reference year is one of the most traditional technique used for the net-gross conversion of income variables. However, Istat decided to setup a new methodology based on the contemporary use of the Siena microsimulation model (SM2-EuSilc) and of a record linkage between survey and administrative data from multiple sources. The available administrative data in terms of net incomes, tax credits and income deductions have been utilized together with survey data as inputs for the SM2-EuSilc model. Administrative data have also been used when appropriate as benchmarks for the microsimulation results. In fact, administrative data and microsimulation estimates are jointly considered for reciprocal comparison and validation and for the construction of the final data set of gross incomes at the individual and household levels. Some significant outputs are finally compared and validated with external sources, mainly taken from National Accounts.

2. The record linkage of administrative and survey data for the italian EU-SILC

The Italian SILC team has developed an innovative strategy in the measurement of self-employment incomes since 2004. This strategy consists in a multi-source data collection, based on personal interviews (PAPI-CAPI) and on the record linkage of administrative with survey data. The term record linkage has been used to indicate the bringing together of two or more separately recorded pieces of information concerning a particular individual or family[4]. The commonly way to combine administrative and survey data is by selecting an individual matching-key able to link the same unit among different data-sources. In other words, the integration of administrative and survey data at micro level is performed by linking individuals through common key variables. The aim of combining administrative and survey data is to improve data quality on income components (target variables) and relative earners by means of imputation of item non-responses and reduction of measurement errors. In addiction matching tax returns records with survey data also provides information at micro level on social security contributions, taxable incomes and tax liabilities. This information is used to measure the gross/net taxable income and represent the input for the SM2 microsimulation model. For the first EU-SILC edition (2004), the integration process involved only two income components: self-employment income and pensions. The following editions also include an integration procedure of information on employment incomes in the tax and survey data sources.

The target population of the EU-SILC survey is the Italian resident population: all private households and their current members residing in Italy at time of data collection. Persons living in collective households and in institutions are excluded from the sample.

The analysis units are adult members (aged 16 and more) living in private households[5].

2.1 The measurement of income components

With regards to the measurement of self-employment incomes in household surveys there are two clear-cut statements, taken from the "Canberra Handbook", that depict the state of the art: "Income data for the self-employed are also generally regarded as unreliable as a guide to living standards"; "Household surveys are notoriously bad at measuring income from capital and self-employment income".

The alternative sources of microdata on earnings from self-employment may not contain the variable 'disposable income'. Survey data may be affected by under-reporting. On the other hand, administrative data gathering individual tax returns do not take account of illegal tax evasion and may not display all the authorized deductions allowed in the calculation of taxable income (tax avoidance). In general, neither taxable income is identical to gross income, nor net taxable income is identical to disposable income. In principle, if the deductions from profits are available to the company owners for their personal use, then they should be considered as components of both the gross and the disposable personal incomes. However, not all the tax abatements allowed are explicitly shown in the tax returns. By definition, tax evasion is also not available in the tax files.

In the EU-SILC project, the standard procedure to measure net self-employment income requires to collect "the amount of money drawn out of self-employment business" only when the profit/loss from accounting books or the taxable self-employment income (net of corresponding taxes) are not available. For the Italian EU-SILC, both tax and survey microdata are available, through an exact matching of administrative and survey records. However, both sources may be affected by under-estimation of self-employment incomes.

[4] Newcombe 1995.
[5] Until 2010, in IT-SILC survey were also interviewed people aged fifteen-year-old.

Moreover, some individuals report self-employment incomes in only one data source. This is the case of some individuals whose professional status at the time of the interview is different from that of the income reference period and of many recipients of small and/or secondary self-employment incomes[6].

Regarding the measurement of income from pensions it is assumed that the administrative data provide more accurate information respect to the survey data. The latter data source is used only if it is impossible to match the sample units to those contained in the Personal Tax Annual Register (unmatched units).

The integration of the administrative sources on pensions and pensioners needs an harmonization of units, definitions and variables and the reconciliation of the incoherencies between the income amounts reported in the different sources. Table 2.1 reports the most relevant meta-information on pension for each administrative data-source. It is noticeable that in most cases it is possible to estimate the final EU-SILC target variables only by bringing together two or more separate pieces of information recorded in different sources.

For example, in order to reckon the net amount received by the elderly separately for each different category of pensions included in the list of target variables, both the "yearly net taxable income from pensions" (Tax Register) and the "monthly gross payments" (Italian Social Security Agency) have to be broken down by kind of pension (employment, early retirement, survivors) and, moreover, to be consistent with the answers given by the respondents, when these are reliable.

The Pension Register collects information at the individual level on the relative beneficiaries, the monthly amount before tax, the classification according to EU-SILC target variables. On the other hand, the Tax Registers record the information on yearly gross/net incomes received by each pensioner without any distinction between the different categories of pensions and is not necessarily consistent with the target variables, namely when a particular kind of pension is tax exempt. In order to join the information of the Tax Registers with the Pension Register we need to define a "harmonized definition of pension income" that is comparable between all these data sources and the EU-SILC project. The common base for the comparison is represented by the "taxable income relating to pensions"[7].

The measurement of employee income is based on the comparison of administrative and survey data on wages and salaries after retention of taxes on labour and mandatory social security contributions at source. The main administrative data source for this income component is represented by the CUD/770 tax statements register. In Italy the employers, as withholding agents, are obliged to declare the net amounts of wages/salaries and of taxes and social contributions annually paid to and for their employees. However, the items included in employee income considered by the administrative source are not exactly the same of the target variable PY010N/G (employee cash or near cash income), therefore it is necessary to reallocate some of them in a proper way.

The administrative net income is obtained as net taxable employee income less taxes and social contributions retained at source. This aggregate is thus compared with the net employee income reported by the respondents in the questionnaire.

---

[6] For a more detailed analysis of this subject it is advised to see Consolini et al. (2006) and Di Marco M. (2006).
[7] See, for more details, Consolini P. (2008).

Table 2.1 – Meta information on pensions/pensioners by administrative sources

Data sources	Variables						Domains	Units
	Gross Income for pension		Net Income for pension		Number of payments	Pension type (Function)		
	Monthly	Yearly	Monthly	Yearly				
Pension Register (PR)	✓(a)	✓(c)	-	-	✓(c)	✓(a)	Census of pensioners of the Italian Social Security System	Pensioner and/or Pension
CUD/770 Tax Register	✓(b)	✓(a)	✓(b)	✓	✓(b)	-	All beneficiaries of taxable pensions	Pensioner
730 Tax Register	✓(b)	✓(a)	✓(b)	✓	✓(b)	-	All beneficiaries of taxable pensions (only.730 Tax Register)	Pensioner
Unico Tax Register	✓(b)	✓(a)	✓(b)	✓	✓(b)	-	All beneficiaries taxable pensions (only.Unico Tax Register)	Pensioner

(a): recorded data
(b): variables derived from the integration of data by different sources.
(c): partially estimated (new pensioners from Pension Registers 2003-2004). In the Pension Registers 2005 data are recorded.

EU-SILC also collects information on several “non-pension cash benefits” by using administrative data sources. In particular, unemployment benefits and family allowances are gathered - on a micro level - from the Inps (National Social Security Institute in Italy) database: “Employees' temporary benefits (GPT) of private sector”. In order to improve the quality on non-pension cash benefits data (i.e. maternity leave, paid sick leave, etc.) new Social security’s databases will be exploited in the next years.

The information on income from capital assets is collected by interviews and the final estimation of this component is typically underestimated as usually happens in income surveys. It is well-known that obtaining accurate and unbiased information on assets income or financial assets is problematic due to the reluctance of the extremely wealthy households to participate in social surveys at all and to respondents’ reticence to declare the ownership of a specific asset. Currently, no administrative data are available to estimate income from capital assets or to adjust the underestimation on financial assets and related incomes.

2.2 The integration methodology

In order to carry out the integration of different databases, some basic requirements have to be satisfied by all sources involved. Namely, the statistical units must be uniformly defined in all sources (harmonisation of units), all sources should cover the same target population (completion of populations), all variables have to be defined and classified in the same way in the different sources (comparability of variables and classifications), all data should refer to the same period or the same point in time[8]. In short, administrative data must be comparable with the EU-SILC survey data.

The technique used to link the administrative units to those in the survey sample is the exact record linkage. This results in combining information related to the same statistical unit by means of identifiers called "matching keys" to obtain an integrated set of information that is exact, in the sense that it actually refers to the matched individual, provided that each statistical unit is associated with a unique identifier not affected by errors. Different typologies of exact record linkage exist: we have used the simplest "one-to-one" relationship, where every statistical unit of a data source is associated with at most one record from the other data source[9]. Records in different data sources have been matched using the Personal Tax Number. Once that is accomplished, the identification numbers are dropped and replaced with an internal anonymous code, according to the policy of the Italian National Statistical Institute.

The integration process between survey and administrative data at the micro level can be summarized in the following three phases[10] (see also Figure 2.1):

a) Input data: the administrative archives

Tax Agency data and Social Security (Inps) data are the administrative data sources involved in the matching process. Personal tax numbers are checked and corrected and the information coming from multiple records relating to the same person is rearranged in order to avoid duplications. In practice this step consists in reading, checking and arranging the tax records' content on the three principal income components: employee income, self-employment income and pensions. At this stage, four relevant sources of microdata have been uploaded: 1) the "Pensions Register (PR)" from INPS (Italian National Social Security Agency); 2) the "CUD/770" tax statements register (of employees, temporary workers and pensioners) from National Tax Agency; 3) the "730" tax returns register from National Tax Agency (taxpayers with at least one CUD/770 tax statement), 4) the "Unico persone fisiche (UPF)" tax returns register of "self-employed" from the National Tax Agency.

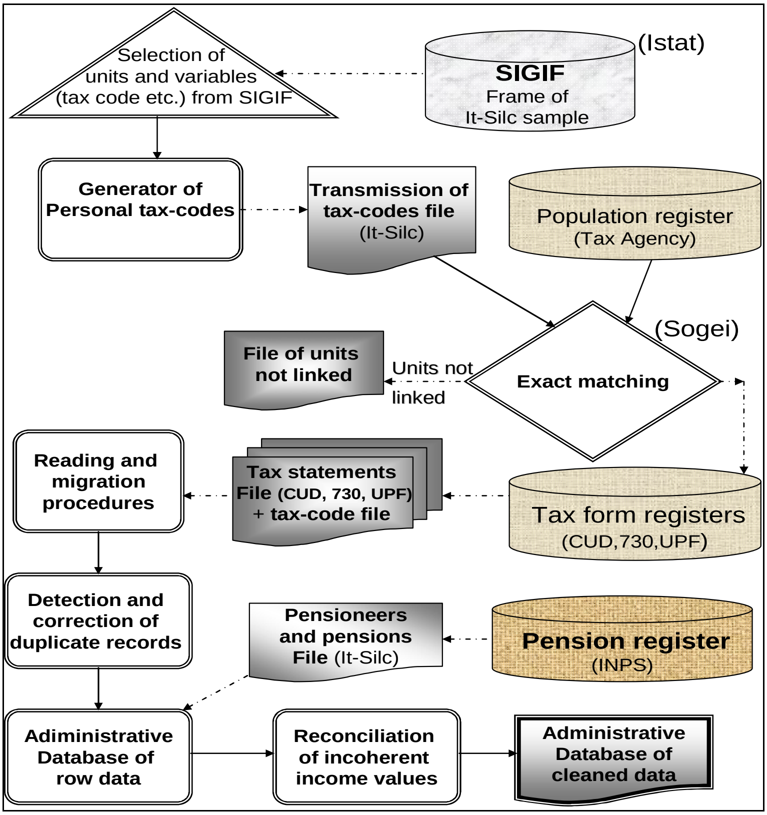
b) The exact matching procedure

At this step the survey and the administrative sources are matched using the Personal tax code number as the key variable. Each sample person is identified with her/his tax code (i.e. the personal identification number assigned to each individual by the Italian tax authorities). The output is a file (matched file) containing information on incomes both from the survey and the administrative archives. More precisely, linkage focuses mainly on adults (15 years and over) that actually participated in the survey. In 2008 the rate of successfully matched records was 96.4%. In other words, the tax source covers 96.4% of the adults interviewed for IT-SILC survey. The unmatched units (3.6%) are either individuals with no tax code available in the Population Registers (2.2%) or persons not included in the initial survey frame but registered later as additional household members by

[8] van der Laan, 2000.
[9] See Newcombe (1988), Herzog, Scheuren and Winkler (2007).
[10] See Consolini P. (2009).

the interviewers (1.4%).

Figure 2.1 - A simplified sketch of data integration process in IT-SILC



c) Detecting and solving incoherencies on income in the matched file

Sometimes the survey and the administrative data sources classify the same income of a recipient under different names. A complex system of editing rules has been established in order to choose which income component must be attributed. Similarly, analysing the coherence between administrative and survey data on the amounts of incomes that go under the same name has required a detailed editing procedure for reconciling monetary values.

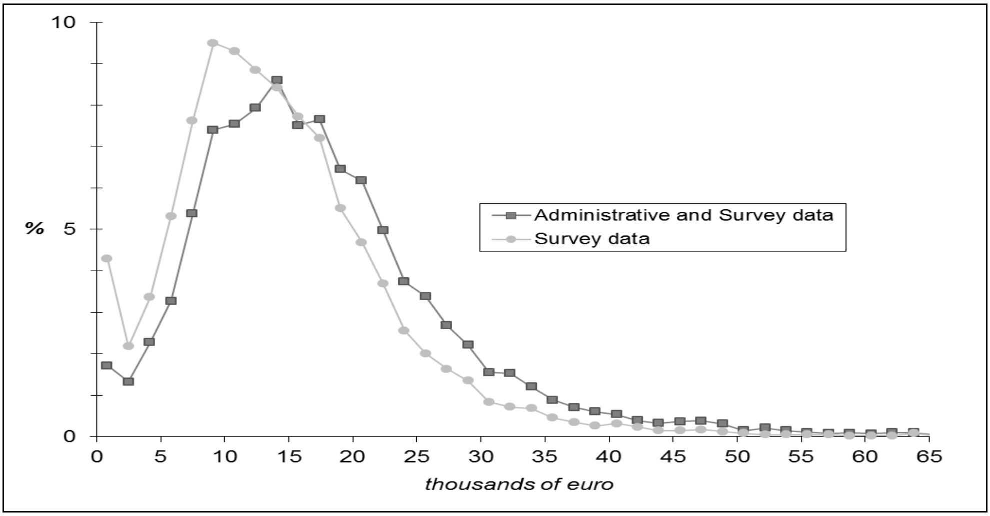
The assumption underlying the fourth step has been that true disposable self-employment income may be under-reported by both sources. In order to minimise under-estimation, self-employment income has been set to the maximum value between the net income resulting from the tax source and the net income reported in the survey. In most cases, comparisons of self-employment income reported in the two sources have been made at the individual level. However, for small family-run businesses, comparisons have been made at the household level, that is by comparing the sums of the self-employment incomes received by all household members in the two sources.

As regards the pensions, when the gross taxable pensions of the Pension Register is compared with the gross income pensions of the CUD/770 tax source, it turns out that for the 84.2% of the matched cases the relative difference between the two amounts is lower than 5%. The assumption underlying the building of the gross/net income from pensions is that the gross income reported in the Pension register is true and the proper information on the tax at source, as well as on tax credits, is included in the other administrative sources. Survey data on pensions (after retention at source) are taken in to account where it is impossible to link the administrative data.

With respect to employment income, we assume that true disposable employee income is included in the administrative source providing that employee does not receive exempt income items (like tips or bonuses) or is employed in sectors of hidden economy (like agricultural, private educational institutes, etc). The CUD/770 tax register includes 99.1% of employee income records reported in all administrative sources.

An assessment of the impact of the multi-source versus survey approach (income data collected by interview) on the equivalent income distribution has been carried out for IT-SILC 2011 edition. As displayed in Figure 2.2, the effect of the inclusion of administrative data involves a shift forward of the income curve. At first glance it seems that the adjustments produce a steady rise in the income levels across the whole survey distribution.

Figure 2.2 – Equivalised income (YEQ) distribution from survey and integrated database (Year 2011)



Merging administrative and survey data definitely brings about a rise of 28.3% in the number of recipients and an increase of 8.8% in the average of self-employment income compared to the exclusive use of survey data (Eusilc 2008 edition). When both sources report information on self-employment incomes, there is some evidence of a higher under-estimation rate on the tax data compared to the survey data. As results from data integration in Eusilc 2008, the number of employee income receivers increases of about 11% whereas employee income increases by about 0.7%.

3. The building of EU-SILC gross income variables

For the estimation of EU-SILC gross incomes, Istat fitted the University of Siena's SM2 microsimulation model to the non-standard case of an integrated dataset from multiple sources. The Siena Micro-Simulation Model (SM2) has been adopted as a recommended procedure by the European Commission for the provision of EU-SILC gross income statistics. The first release of SM2 had been developed by Siena University team in 2003 and initially applied to the ECHP (*European Community Household Panel*) survey data. For the EU-SILC project, SM2 has been updated and extended to become a general and flexible tool for the "net to gross" and "gross to net" conversion of income variables that can simulate the functioning of the tax-benefit system of different countries[11]. In fact, the model can be applied to diverse input data collected in various forms across and within countries and it is able to generate variables in a comparable and standardised 'multi-country' format.

The model estimates income by component, breaking down gross amounts into taxes, social insurance contributions, social transfers, and net/disposable incomes. All the information on income components have to be collected, compiled or imputed in some form, and the model converts it, under a specified national tax system, to the standard form required by the EU-SILC project.

The SM2 consists of a standardised set of routines, which can deal with a great diversity of input data forms and national tax systems. Country-specific routines are required to convert the input data formats and to define the parameters of the national tax system in an appropriately standardised form. These routines are taken as inputs by the core of the model, that generates the required standardised outputs. The system maintains a clear distinction between the general and the country-specific parts, and it is developed to maximise the part which can be standardised in order to be easily applied for different tax benefit systems [12].

In 2004 Istat decided to test the application of the SM2 and use both administrative archives and sample survey data for the net-gross conversion of EU-SILC variables.

The data production process of the EU-SILC gross income variables can be summed up in three important steps: the first one is the adaptation of the model SM2-EuSilc from SM2; the second one is the integration of survey data and administrative data used jointly with microsimulations and the third one is the setup of the final dataset of individual and household income target variables.

The implementation by Istat of the SM2-EuSilc model has required the transition from the preliminary version applied to the ECHP data to the version applied to the EU-SILC data and the construction of the input and auxiliary variables on the basis of information

[11] The model was set up under the Eurostat project " Development of Appropriate Modelling or Imputation to Construct the EU-SILC Target Income Variables for Each EU Member States".

[12] For a detailed description of the model see: Betti et al, 2011.

collected by the new survey. Originally, the SM2 input file was based on the Eurostat releases of the ECHP User and Producer databases for three countries (Italy, France, Spain, see Eurostat, 2004). As regards to Italy, the income reference year was 1998 and the tax rules were those of the year 2003, in order to include the then recent tax reform.

The adaptation of the model to the new EU-SILC survey called for new procedures for the setup of the input file and implied the adjustment of some conversion routines of SM2.

The first step in the construction of the input file was a direct substitution, where possible, of the Echp variables with the new ones. The second step was the construction of the auxiliary variables based on the information available in the new survey. Several auxiliary variables were required for the input file of SM2 and particular attention was paid to the construction of the tax units. To identify the tax units at the household level, the "family procedure" used in Istat social surveys was applied. The procedure consents to classify the households on the basis of the couple and parental relationships, identifying the dependent persons and those entitled to the family tax credits.

Compared to ECHP, the new EU-SILC survey collects detailed information on sector of activity, work status, number of months in a given status and firm size: information that is useful for calculating the social security contributions for dependent and self-employed workers. Moreover, the breakdown of sickness and invalidity benefits is available in EU-SILC as well as data on pension contributions made to private insurance companies, which could be deducted from the tax base of the Italian personal income tax.

The transition to the new survey required also an adjustment of some conversion routines of SM2, in particular for the calculation of self-employment income and the estimation of the IRAP tax (regional tax on productive activities) paid by the self-employed, including the self-employed. In the ECHP survey, self-employment income was in fact collected as a gross amount, while in EU-SILC it is recorded as net income.

Additional modifications in SM2 conversion procedures were needed for the calculation of the income of the Co.Co.Co. (temporary subcontractors) which is nominally included in self-employment income, but in fact is treated as employment income. Data on this kind of workers were not available in the ECHP survey, and a variable defining the likelihood for an employee to work under a Co.Co.Co contract was estimated in SM2, using external sources. Extra amendments of SM2 procedures were needed also for the estimation of family deduction for dependent persons in order to include the second module of the IRPEF (personal income tax) reform of 2005 and for subsequent amendments to tax legislation occurred in 2007 and 2009.

3.1 A description of the Italian tax system as integrated in the model SM2-EU-SILC

The main components of the Italian tax system are summarised in table 3.1, which displays which income components are liable to social insurance contributions and income tax, respectively. Employment and self-employment incomes are subject to social insurance contributions, determined as a function of gross income (Gi), and to income taxation. Social insurance contributions are withdrawn from gross income to obtain the gross taxable income, as they are not subject to the main Italian personal income tax (Irpef).

Irpef is calculated by applying marginal progressive tax rates to increasing income

brackets and for this purpose the incomes subject to Irpef are pooled together. Specifically self-employment income is also subject to a special tax, Irap (Tax on income from production activities), determined as a function of value added, that includes gross taxable income from self-employment. This kind of 'double taxation' at a flat rate is handled in the model by simply treating it as a 'negative tax credit'.
In fact, a distinctive trait of the model is that properly defining certain ‘special deductions or tax credits’, in addition to the typical tax benefits, many complexities of a tax regimes can be incorporated into the standardised procedures. Income components which are not subject to the Irpef are automatically removed from the common pool by just specifying their 'component-specific deductions' as equal to the component's total gross taxable income (in order to exclude their contribution to net taxable income). This applies for example to tax-exempt benefits. Moreover if a component is taxed at a flat rate separately from the pool, it is possible to specify its ‘special deduction’ as equal to the component's total gross taxable income in order to remove it from the pool, and its ‘special tax credit’ as a negative quantity. In this way, the component taxed at a flat rate makes no contribution to the tax liability of the pool, but the final tax liability is automatically increased by the appropriate amount. Tax on property assets or financial capital income can be handled in a similar way.

Table 3.1 - Main components of income, tax and social insurance contributions in the Italian fiscal system (year 2011)

N	Income components	Social Insurance Contributions (Si)	Tax	Included in common pool	Component specific	
					Deduction (Di)	Tax Credits (Ci)
1	Employment income	Employer's $S_0(G_1)$ Employee's $S_1(G_1)$	IRPEF	X		$C_1(Y_1)$
2	Self-employment income	$S_2(G_2)$	IRPEF	X		$C_2(Y_2)$ $-f_2(H_2)$ "IRAP" (a)
3	Pensions		IRPEF	X		$C_3(Y_3)$
4	Property (rental and cadastral) income		IRPEF (b)	X		
5	Financial Capital income		Taxed at source (flat rate K_6)		H_6	$- K_6 \times H_6$
6	Education related benefits, Unemployment benefits		IRPEF	X		
7	Family benefits, Sickness invalidity benefits (c), Housing allowances, Any other personal benefits		Tax exempt		H_7	
8	Property value		IMU (on value of real estate)			-f8(value)

(a) Irap: Tax on income from production activities. f(..) stands for “a function of”.
(b) On total cadastral and on 85% of the rental income, if not subject to the new regime of rental income flat rate “cedolare secca”, launched in 2011.
(c) Part of the benefits are taxable.

In Italy, the incidence of social insurance contributions on income from work is different according to the source of income, occupational status and sector of activity. Employers' and employees' social insurance contributions are imposed on gross earnings from wages. For dependent workers there are minimal and maximal amounts of contributions to be paid. These two limits depend on firm size (number of workers), sector of activity (based on the classification NACE Rev.2) and occupational status (workers, employees, executives).

Self-employed workers' social insurance contributions are divided into three main categories: for general self-employed (i.e. craftsmen or workers in commerce), agricultural self-employed, and professional persons. The social insurance contribution rates are different in these categories and apply to income brackets and they depend also on the age of the worker. There is also a common minimum and maximum base of contribution for general self-employed (in 2011, euro 14.552 and euro 71.737) and if the self-employment income is under the minimum, they have to pay as the minimum. The agricultural self-employed have to pay a fixed amount depending on their annual income brackets.The self-employed professional persons include partners in a company, and professional workers (entrepreneur or owner, assistant of a household firm) divided in two different categories: (a) professional persons not registered in any other compulsory social insurance institution[13] and (b) professional registered in any other compulsory social insurance institution who pay supplementary contributions, as well as the occasional self-employed workers, if their annual gross income is exceeding 5.000 euro. The first category also includes the PhDs or research grant recipients and the CoCoCo (temporary subcontractors) with a special status in employment that is essentially intermediate between dependent and independent employment. The CoCoCo are essentially considered self-employed, but they have particular treatment in the Italian fiscal system and their income is treated as employee's income and, for this reason, the social insurance contributions are also paid by the employer. These contributions are, however, lower than the normal ones. For the first category the social insurance contributions rate, in 2011, account for 26,72 per cent of the annual self-employment income. In the model this rate is applied for those professional persons who only have this type of income, without any other kind of incomes or pensions. For the other professional persons the 17 per cent of annual gross income is applied by the model.

3.2 The production of EU-SILC gross income statistics

The statistical production process of the EU-SILC income variables is made up of several complex phases that can be summarized in two broad sequential steps: first the construction of net income and then the production of the income before taxes and social insurance contributions. The availability of data from administrative sources, used from the stage of construction of net incomes, has enabled the joint, innovative use of the microsimulation model and administrative archives. The integration of survey data and register data in EU-SILC has the most important aim to reduce the under-estimation of incomes on the basis of available information (survey and registers). As is well known, data

[13] Professional persons registered in the compulsory social insurance institution "Gestione separata" of the Italian National Institute of Social Security.

from income tax returns could not contain information on a number of income components (untaxed incomes, incomes taxed separately or subject to withdrawal taxes) and may have problems of coverage in relation to the individuals included in a sample survey. The survey data, in turn, may be subject to withholding of information (reticence), under-reporting or inadequate representativeness of certain types of income or income recipients. The joint use of survey and administrative data enhances the advantages obtainable from the exclusive use of fiscal archives on the one hand and of microsimulation techniques on the other.

For the construction of the gross incomes variables, the “730 tax returns” and the “UPF tax returns” provide data on net and gross incomes, taxes at national and regional level, and data on tax credits[14], income deductions[15] of declarant and spouse[16]. It is worth noting that in any microsimulation model, as in the previous SM2, the income deductions and tax credits based on consumption expenditures usually needed to be estimated by regression technique based on external sources. Respect to what done in the phase of construction of net income a new integrated data set is then made with data on taxes, income deductions and tax credits. Before using all the available information (fiscal and survey data) as input file of the model SM2-EuSilc an additional procedure for checking the consistency and accuracy of the administrative data is applied. In this way a number of anomalies between withholding taxes, taxes paid, social security contributions and corresponding incomes are eliminated. Finally the SM2-EuSilc outputs are compared with the available administrative gross figures at the micro level in order to assess the quality of microsimulation estimates and for reciprocal validation.

The final database of individual and household incomes gross of tax and social security contributions is therefore constructed as the sum of net incomes, taxes paid and withholding taxes from administrative sources, if available, or as the sum of net incomes and microsimulated taxes[17]. It includes, additionally, social security contributions paid by workers and employers estimated by SM2-EuSilc. In fact, the available registers on compulsory social insurance contributions only cover data on employees of private sector (not employers) and on employees and employers of public sector and there is merely a partial coverage of the social insurance contributions of self-employed drawn on the UPF tax returns.

It should be noted that the implemented methodology is essentially based on a strategy of combining fiscal data and microsimulation estimates. In more details it can be said that when the net administrative incomes are higher than the survey incomes, the EU-SILC net and gross incomes completely arise from administrative data, while the social insurance

[14] Tax credits for expenditures as per Section I and III of Part RP of UPF 2012 (medical expenses, vehicle expenses and guide dogs for disabled people, mortgage interest, life assurance and accident insurance, tuition fees, funeral expenses, care expenses, expenses for children’s sports activities, estate agents’ fees, rent costs for students living away from home, other costs and expenses for building renovation work which are deductible at the rate of 41 per cent or 36 per cent) and other tax credits as per Section IV, Section V, Section VI and Section VII of Part RP.

[15] Deductions for principal dwelling and deduction for expenditures as per Section II of Part RP of UPF 2012 (social security and welfare contributions, regular maintenance allowances paid to spouse, social security contributions for home helps and cares, donations to religious institutions, medical and care expenses for disabled persons, supplementary health insurance and other expenditures).

[16] For taxpayers who filed both 730 and UPF returns, the UPF form was used as it generally contains additional, subsequent information compared with the 730 form.

[17] A stochastic component has been added to the withholding taxes and taxes paid from administrative sources to render the information used anonymous.

contributions are estimated by the model. Consequently the final EU-SILC gross variables do not differ from the fiscal ones. On the opposite, when the survey incomes are higher than the administrative data, the net incomes are those taken from the survey (collected or imputed), while the taxes derived from administrative sources, since these are taxes actually paid by the taxpayers. As always the social insurance contributions are estimated by SM2-EU-SILC. In such a case the final EU-SILC gross variables are essentially different from the fiscal gross data. It is worth mentioning that when the surveyed incomes are higher than the register data, the difference between the surveyed data and the tax data could not be considered as a direct measure of illegal tax evasion. As a matter of fact it is not possible for us to distinguish between the legal tax avoidance, allowed by the national fiscal system, and the real tax evasion[18].

As a direct result of the applied methodology the typical ‘adjustment factors’ used in any microsimulation model for correcting the disposable income and the gross income values in order to take into account the tax evasion are not applied. In effect EU-SILC disposable income partly includes income not reported to tax authorities, while the taxes for the most part are those derived from the income tax returns and do not require any adjustments.

Table 3.2 - IT-SILC target variables net-gross ratio and gross and net distribution by income components - Year 2011 (percentage values)

VARIABLES 2012 (INCOME REFERENCE YEAR 2011)	Ratio Net/Gross	Distribution	
		Gross	Net
Income from work	**71.4**	**66.5**	**63.1**
PY010 Employee cash or near cash income	72.1	47.6	45.6
PY050 Cash benefits or losses from self-employment	69.5	18.9	17.5
Property income	**73.0**	**4.0**	**3.9**
HY090 Interest, dividends, profit from capital investments in unincorporated business	79.4	1.1	1.2
HY040 Income from rental of a property or land	70.4	2.9	2.7
Taxable benefits	**83.4**	**28.2**	**31.2**
PY090 Unemployment benefits	86.0	2.1	2.5
PY100 Old-age benefits	82.5	24.1	26.4
PY110 Survivor’ benefits	85.4	0.8	0.9
PY130 Disability benefits	96.4	1.1	1.4
Tax-exempt social transfers	**100.0**	**1.3**	**1.8**
PY140 Education-related allowances	100.0	0.1	0.2
HY050 Family related allowances	100.0	0.6	0.8
HY060 Social assistance	100.0	0.1	0.1
HY070 Housing allowances	100.0	0.0	0.0
HY080 Regular inter-household cash transfer received	100.0	0.4	0.6
Total	**75.2**	**100.0**	**100.0**

As shown in the table 3.2, the net/gross ratio varies by component for the differences in component-specific deductions and tax credits, and also in the social insurance contributions. The net-to-gross ratio is lower for income from work (71.4%) than for the other components, due to the social insurance contributions to which such income is

[18] In literature there are several works on tax evasion based on such differences. This subject actually goes beyond the present chapter focused on the Eu-Silc production process utilized as input file of the ISTAT microsimulation model.

subject. The ratio of net to gross taxable income of other incomes varies approximately from the low of 70.4% for property income, to 83.4% for various taxable benefits, to of course 100% for social assistance, housing and other tax-exempt benefits. The distribution of gross income shows clearly that the main income component is represented by income from work (66.5%), followed by old age benefits (24.1%). The differences in gross and net distribution proves that the tax burden is higher in income from work than the other components, like taxable benefits and property income.

In the following tables the comparison between EU-SILC and some appropriate external sources are also presented. Data from National Accounts, Labour Force Survey by Istat and data from Fiscal Agencies of the Ministry of the Economy and Finance and the Pensions Register by INPS (National Institute for Social Security) are used as external benchmarks.

The comparison of EU-SILC data with National Accounts figures are shown in table 3.3. The table reports the breakdown of total gross income into social insurance, tax and net components. EU-SILC estimates embrace all income components of target variables even those not included at present in the total gross household income (HY010) (i.e. imputed rent, all fringe benefits, own consumption, employers' social insurance contributions, interest paid on mortgage). On the average, net income, after tax and social insurance contributions including employers' contributions, accounts for 68.3% of total gross. The agreement of EU-SILC results and National Accounts figures is good and let the EU-SILC results satisfactory.

Nonetheless some aspects have to be considered when comparing EU-SILC with National Accounts. NA estimates generally use all the administrative data sources which are integrated in EU-SILC and as it is well known NA estimates are adjusted to account for the grey economy. However the grey economy is partially covered in EU-SILC given that some interviewees report income that are not enclosed in tax registers, including both tax avoidance/evasion and tax exempt. On the one hand EU-SILC integration methodology applies the rule of the maximum between survey and administrative income level, consequently the mean income of EU-SILC is usually higher than the administrative one (which is employed in NA estimates). Moreover EU-SILC survey, as well as other income surveys, typically under-estimates financial capital incomes, which are subject to tax withholding at source at some flat rate, whereas EU-SILC estimation method of imputed rent produces higher value than NA aggregate. Finally it is expected that the combined effect of the above mentioned features explains the closeness between the two data sources estimates.

Table 3.4 shows the comparison of EU-SILC income recipients and some external sources. The EU-SILC number of employees who have received wage or salary positively approximates the number of income earners from National Fiscal Agency data, which represents the universe of taxable employee income recipients. Differences in applied definitions (i.e. domestic vs resident employment), reference period and coverage of the two data sources can well explain the discrepancies in estimates. Furthermore the tax register does not report information on incomes and employees of the hidden economy that, as stated before, are partially included in the survey.

For lack of harmonization and divergence in definition of self-employment income, National Accounts are not directly comparable with EU-SILC estimates and other sources are employed. It should be noted that important differences in definition of self-employed make the agreement reasonable but not excellent. In fact in LFS a worker is classified as an independent on the basis of his/her main activity and in NA the estimate of self-employed units is in term of full time equalised workers. On the other hand the EU-SILC estimate is referred to the number of people whose earnings from self-employment may have been temporary and/or from a secondary working activity. Data on beneficiaries for three kind of functions - old-age, survival and disability (according to ESSPROS classification) – are also reported and the comparison with external sources shows that EU-SILC estimates are quite close to the administrative data.

Table 3.3 - Distribution of total gross income – Year 2011 (euro per capita and percentage values)

	IT-SILC 2012 (income reference year 2011)		Istat N.A. 2011	Difference (% point)
Gross including SI	21697	100	100	
SI contributions	3917	18.1	17.3	0.8
- Employers' contribution	2718	12.5	12.2	0.3
- Employees' contribution	711	3.3	2.9	0.4
- Self -employment contribution	488	2.2	2.2	0.1
Gross taxable	17779	81.9	82.7	-0.8
Personal income tax and financial tax	2964	13.7	13.3	0.4
Net income	14814	68.3	69.4	-1.1

Sources: Istat (2012) and Istat (2013)

Table 3.4 - Comparison of It-Silc income recipients and some external sources - Year 2011 (Thousands of units)

	It-Silc	Fiscal Agencies	National Accounts (Ula)(a)	Labour Force Survey (LFS)	Pension Register of INPS (b)	Differen-ce (%)	Differen-ce (%)
Persons who have received wage and salary (cash or near cash)	21,459	20,951				2.4	
Persons who have received cash benefit or losses from self-employment	7,812		6,712	5,727		16.4	36.4
Beneficiaries of Old-age-Survival-Disability pensions	16,268				15,998	1.7	

(a) Full time equivalent unit of workers.

(b) Severance recipients and persons aged under 15 and/or residing abroad are not included.

Sources: Istat (2012) and MEF (2013)

4. Concluding remarks

This paper provides a thorough review of the methods used for the estimation of net and gross income variables of EU-SILC survey and for enhancing the quality of income statistics that are produced. The methodology is essentially based on the integration of survey data and administrative datasets at a micro level. The Italian mix-mode collection of data on income (EU-SILC) represents the first example in the field of integration of survey and administrative data among European NSIs. Producing statistics from administrative sources means that data collection, editing and other kinds of data processing are done by methods different from the traditional ones. Instead of making quality controls of data received from the individuals, in this new perspective it is necessary to modify administrative data on the basis of our knowledge of differences in definitions and coverage between the administrative sources and the statistical needs. The administrative data are essentially exploited in order to fill in the survey missing values, correct outliers or unreliable values and produce the income variables, improving the quality of the final estimates. The joint use of survey data and administrative archives also allowed to use a more innovative methodology than the traditional microsimulation model for producing incomes before taxes and social insurance contributions. The double use of the tax sources on one side and the microsimulation estimates on the other enhances the advantages obtainable with the exclusive use of one of the two instruments. Administrative data especially in terms of income deductions and tax credits are used in the input file of the model instead of estimation by regression technique based on external sources. Moreover the linkage with administrative data has the advantage of reciprocal comparison and validation of the final estimates.

The achieved results of our process of integration allow to conclude that the best way for collecting household’s income in Italy is to combine administrative data and survey income data. In effect by means of integration is definitely possible to improve the coverage and the quality of income data. Several projects for further improving data quality are conducting at ISTAT aiming at extending the administrative sources used and the timeliness of data. In a short time it will be a more massive use of administrative data in order to replace information from survey questions (i.e. pensions), while there is a transition towards a multi-mode data collection based on computer assisted techniques such as CATI (Computer Assisted Telephone Interviewing) or even CAWI (Computer Assisted Web Interviewing). The aim is finally to have shorter questionnaires and decrease data collection costs, reduce response burden, and achieve a better measurement of income data.

References

Betti G., G. Donatiello and V. Verma. 2011. The Siena Microsimulation Model (SM2) For net-gross conversion of Eu-Silc income variables. *The International Journal of Microsimulation.* 4(1): 35-53. http://www.microsimulation.org/IJM/IJM_V4_1.htm.

Burricand C. 2013. "Transition from survey data to registers in the French SILC survey", Eurostat. *The use of registers in the context of EU–SILC: challenges and opportunities*. Statistical Working Paper. Luxembourg. Publications Office of the European Union 2013. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-TC-13-004/EN/KS-TC-13-004-EN.PDF

Consolini, P., M. Di Marco, R. Ricci, R. and S. Vitaletti. 2006. "Administrative and Survey Microdata on Self-Employment: the Italian Experience with the Eu-Silc project". Iariw 29th General Conference. Joensuu. Finland, 20-26 August.

Consolini P. and G. Donatiello. 2013. "Improvements of data quality through the combined use of survey and administrative sources and micro simulation model". Eurostat. *The use of registers in the context of EU–SILC: challenges and opportunities*. Statistical Working Paper. Luxembourg. Publications Office of the European Union 2013. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-TC-13-004/EN/KS-TC-13-004-EN.PDF

Consolini, P. 2007. "Experiences on the harmonization of the definitions, the variables and the units for the Eu-Silc project in Italy". Working package WP2: Recommendations on the use of methodologies for the integration of surveys and administrative data. Final report of "ESSnet Statistical Methodology Project on the Area: Integration of survey and administrative data". pp.58-67. http://cenex-isad.istat.it/dokeos/document/document.php

Consolini P. 2009. *Integrazione dei dati campionari Eu-Silc con dati di fonte amministrativa*. Collana Istat Metodi e Norme vol. 39/2009. Roma. http://www3.istat.it/dati/catalogo/20090318_00/

Di Marco M. 2006. "International Comparability of Microdata on Incomes: Lessons From the Eu-Silc Project". VIII International Meeting on Quantitative Methods for Applied Sciences. Certosa di Pontignano (Siena). 11-13 September.

Donatiello G., G. Betti G., P. Consolini. 2012. *The Construction of Gross Income Variables of Eusilc (Eu Statistics on Income and Living Conditions) in Italy: A Mixed Strategy Using Microsimulation and Administrative Data*. Università degli Studi di Siena. *Quaderni del Dipartimento di Economia Politica e Statistica*. n. 652 – Settembre. http://www.econ-pol.unisi.it/quaderni/652.pdf

Donatiello G. 2011. *La metodologia di stima dei redditi lordi nell'indagine Eu-Silc - Indagine europea sui redditi e le condizioni di vita delle famiglie*. Collana Istat Metodi e Norme vol. 49/2011. Roma.

http://www3.istat.it/dati/catalogo/20110726_00/metodologia_stima_redditi_lordi_indagine_eu_Silc.pdf

Eurostat. 2004. *Income in EU-SILC: Net/Gross/Net conversion. Report on common structure of the model; model description and application to the ECHP data for France, Italy and Spain*, prepared by V. Verma, G. Betti and co-researcher. EU-SILC 133/04, Luxembourg, 2004.

Heuberger R., T. Glaser T., and E. Kafka. 2013. “The use of register data in Austrian SILC survey”. *The use of registers in the context of EU–SILC: challenges and opportunities*. Statistical Working Paper. Luxembourg. Publications Office of the European Union 2013. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-TC-13-004/EN/KS-TC-13-004-EN.PDF.

Herzog, T.N., F.J. Scheuren, , W.E. Winkler. 2007. *Data quality and Record Linkage Techniques*. New York: Springer ed.

Kapteyn A., J. Y. Ypma. 2007. Measurement Error and Misclassification: A Comparison of Survey and Administrative Data. *Journal of Labor Economics*, 25 (3): 513-551. http://www.jstor.org/stable/10.1086/513298

Istat. 2012. *“Condizioni di vita (UDB It-Silc)”*. Roma: Istat. (microdata) http://www.istat.it/it/archivio/4152.

Istat. 2013. “*National Accounts Years 2011-2012”*. Roma. Istat.

MEF, Ministero dell’Economia e delle Finanze (2011). “Dichiarazioni fiscali”. http://www.finanze.gov.it/stat_dbNew2011/index.php

Méndez Martín J.M. 2013. “Reconciliation of income data from survey and from administrative sources”. *The use of registers in the context of EU–SILC: challenges and opportunities*. Statistical Working Paper. Luxembourg. Publications Office of the European Union 2013. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-TC-13-004/EN/KS-TC-13-004-EN.PDF.

Newcombe H.B. 1988. *Handbook of Record Linkage: Methods for Health and Statistical Studies. Administration, and Business*. Oxford (UK): Oxford University Press.

Nordberg L. 2003. “An Analysis of the Effects of Using Interview versus Register Data” in Income Distribution Analysis Based on the Finnish ECHP-surveys in 1996 and 2000, Chintex Working Paper #15, Work Package 5, December 22 2003.

UNECE. 2011. *Canberra Group Handbook on Household Income Statistics.* Second Edition. United Nations. New York and Geneva.

van der Laan, P. 2000. *Integrating administrative registers and household surveys*. Netherlands Official Statistics. 15. Summer 2000.

Wallgren, A. and B. Wallgren. 2007. *Register-based Statistics: Administrative Data for Statistical Purposes*. New York: John Wiley and Sons.

Norme redazionali

La Rivista di statistica ufficiale pubblica contributi originali nella sezione "Temi trattati" ed eventuali discussioni a largo spettro nella sezione "Interventi". Possono essere pubblicati articoli oggetto di comunicazioni a convegni, riportandone il riferimento specifico. Gli articoli devono essere fatti pervenire al Comitato di redazione delle pubblicazioni scientifiche corredati da una nota informativa dell'autore contenente attività, qualifica, indirizzo, recapiti e autorizzazione alla pubblicazione. Ogni articolo prima della pubblicazione dovrà ricevere il parere favorevole di due referenti scelti tra gli esperti dei diversi temi affrontati.

Per l'impaginazione dei lavori gli autori sono tenuti a conformarsi rigorosamente agli standard editoriali fissati dal Comitato di redazione e contenuti nel file RSU stili o nella classe LaTex, entrambi disponibili on line. La lunghezza dei contributi originali per entrambe le sezioni dovrà essere limitata entro le 35 pagine. Una volta che il lavoro abbia superato il vaglio per la pubblicazione, gli autori sono tenuti ad allegare in formato originale tavole e grafici presenti nel contributo, al fine di facilitare l'iter di impaginazione e stampa. Per gli standard da adottare nella stesura della bibliografia si rimanda alle indicazioni presenti nel file on line.

Tutti i lavori devono essere corredati di un sommario nella lingua in cui sono redatti (non più di 120 parole); quelli in italiano dovranno prevedere anche un abstract in inglese.

Nel testo dovrà essere di norma utilizzato il corsivo per quei termini o locuzioni che si vogliano porre in particolare evidenza (non vanno adoperati, per tali scopi, il maiuscolo, la sottolineatura o altro).

Gli articoli pubblicati impegnano esclusivamente gli autori, le opinioni espresse non implicano alcuna responsabilità da parte dell'Istat.

Per contattare la redazione o per inviare lavori: rivista@istat.it. Oppure scrivere a:
Segreteria del Comitato di redazione delle pubblicazioni scientifiche
all'attenzione di Gilda Sonetti
Istat
Via Cesare Balbo, 16
00184 Roma

La Rivista di Statistica Ufficiale accoglie lavori che hanno come oggetto la misurazione e la comprensione dei fenomeni sociali, demografici, economici ed ambientali, la costruzione di sistemi informativi e di indicatori come supporto per le decisioni pubbliche e private, nonché le questioni di natura metodologica, tecnologica e istituzionale connesse ai processi di produzione delle informazioni statistiche e rilevanti ai fini del perseguimento dei fini della statistica ufficiale.

La Rivista di Statistica Ufficiale si propone di promuovere la collaborazione tra il mondo della ricerca scientifica, gli utilizzatori dell'informazione statistica e la statistica ufficiale, al fine di migliorare la qualità e l'analisi dei dati.

La pubblicazione nasce nel 1992 come collana di monografie "Quaderni di Ricerca ISTAT". Nel 1999 la collana viene affidata ad un editore esterno e diviene quadrimestrale con la denominazione "Quaderni di Ricerca - Rivista di Statistica Ufficiale". L'attuale denominazione, "Rivista di Statistica Ufficiale", viene assunta a partire dal n. 1/2006 e l'Istat torna ad essere editore in proprio della pubblicazione.

ISSN 1828-1982

€ 15,00