Developing and Validating Regional Microsimulation Models. TREMOD: The Tax-Benefit Model of the Italian Province of Trento

by Davide Azzolini[*], Martina Bazzoli[**], Silvia De Poli[***], Carlo Fiorio[****] *and* Samuele Poy[*****]

ABSTRACT

A large number of microsimulation models have been developed in Italy over the past years. Only few of them have been created to study specific regional contexts. This paper illustrates the development and validation of TREMOD, a new "static" tax-benefit microsimulation model for the Italian province of Trento. TREMOD is based on the EUROMOD platform and is proposed as an ex-ante evaluation model applicable to local taxation and welfare policies. The main strength of TREMOD is the high quality of the data used. The input database has been obtained by matching data from a local representative survey on households' life conditions with administrative data on individual income tax returns. This aspect is one of the main strengths of TREMOD compared with other experiences in microsimulation modelling. As shown in this paper, the combination of survey and administrative data yields substantial gains in simulation precision.

JEL: H24, I32, I38

Keywords: Fiscal policies, microsimulation model, validation, administrative data

[*] FBK-IRVAPP, Vicolo dalla Piccola, 12, Trento (IT), Phone +39 0461314234. E-mail: azzolini@irvapp.it.

[**] FBK-IRVAPP, Vicolo dalla Piccola, 12, Trento (IT), Phone +39 0461312297. E-mail: bazzoli@irvapp.it.

[***] FBK-IRVAPP, Vicolo dalla Piccola, 12, Trento (IT), Phone +39 0461314222, Corresponding author's. E-mail: depoli@irvapp.it.

[****] FBK-IRVAPP and University of Milano, Via Conservatorio, 7, Milano (IT), Phone +39 0250321540. E-mail: carlo.fiorio@unimi.it.

[*****] Università Cattolica del Sacro Cuore, Largo Gemelli 1, Milano (IT), Phone +39 0272342778. E-mail: samuele.poy@unicatt.it.

The creation of TREMOD has been funded by the Department of Labour and Welfare of the Province of Trento (Italy). We are indebted to the Statistical Office and the Labour Agency of the Province of Trento for their invaluable support. TREMOD has been constructed using EUROMOD version 1.9. We thank the EUROMOD Core-Team for giving us access to EUROMOD and helping us in its use. We are also grateful to Erich Battistin, Daniele Checchi, Livia Ferrario, Enrico Rettore, Antonio Schizzerotto, Ugo Trivellato and two anonymous referees of Economia Pubblica/The Italian Journal of Public Economics, for very useful comments and suggestions. Finally, special thanks go to all the colleagues who kindly helped us in collecting valuable information regarding the Italian simulation models.

1. Introduction

An increasing attention has been paid by stakeholders, researchers and policy makers toward ex-ante evaluation of public interventions. Most part of this public debate concerns tax and welfare policies. In many cases, relevant public actors are interested in predicting the effects of different policy scenarios. As argued by Mitton et al. (2000), microsimulation models represent a valid instrument to assess the effects of policy change using individual and/or household data. Microsimulation models can be broadly classified with respect to their underlying assumptions regarding agents' behaviors (i.e., if they change or not in response to a given policy change), and the time horizon of agents' decisions (be it dynamic, i.e. consisting in an intertemporal budget set, or static, in the case of a permanent budget set). Our contribution refers to the large literature on static (not behavioral) microsimulation models.

Ex-ante impact evaluation can be of particular interest in local/regional contexts. Unfortunately, to this regard available instruments in Italy are few. Our paper tries to fill part of this lack. The scope of this paper is illustrative. Specifically, it: *i)* presents and describes in detail the main features of TREMOD, the static tax-benefit microsimulation model for the Province of Trento; *ii)* discusses the data used by TREMOD and tests its performance driven by the use of a local survey matched with administrative income data.

TREMOD has been developed by *The Research Institute for the Evaluation of Public Policies* of Bruno Kessler Foundation (hereafter "FBK-IRVAPP") by adapting the EUROMOD microsimulation model platform (Sutherland and Figari, 2013). As discussed by Di Nicola et al. (2015), the choice of alternative input data sets for static tax-benefit microsimulation models is a key issue for their performance. In this regard, TREMOD provides a substantial improvement on most of the existing microsimulation

models (for a previous application in the Italian context see Di Nicola et al., 2015) becauseit uses a combination of survey and administrative micro data (instead of just survey data). More precisely, TREMOD uses a combination of microdata from a local representative survey on households' life conditions (*Indagine sulle condizioni di vita delle famiglie trentine,* hereafter also called 'ICFT data') with official administrative data on individual income tax returns obtained from the Italian Revenue Agency (RA). We will show that this combination of data sources allows for an improved precision in the measurement of individual incomes.

The paper is structured as follows. Section 2 presents a comprehensive review of national and regional static microsimulation models developed in Italy. Section 3 provides an overall description of TREMOD, highlighting its analytical potential and distinctive characteristics, as well as illustrating some of its potential practical uses. Particularly, the section describes the database used for the construction of TREMOD and its statistical representativeness. Section 4 assesses the estimation power of TREMOD, comparing some economic variables simulated by the model with the same variables observed in external administrative data sources. Section 5 conducts a comparison between the microsimulation model that we propose (which, as mentioned, integrates survey data and administrative data) with an alternative model based only on survey data (IT-SILC, as in the case of EUROMOD). Finally, Section 6 concludes by summarizing TREMOD's main strengths and envisaging possible developments and applications.

2. Microsimulation models in Italy

Over the past thirty years, several microsimulation models have been developed in Italy. Table 1 lists and shortly describes, to our knowledge, the most relevant static microsimulation models at the national level[1]. We collected detailed information on 14 models regarding the data sources used to recover income information; the frequency of updates; how the model's development has been funded; whether the model's algorithm and data are

[1] The list includes all models on which a scientific publication exists (loosely defined as any academic manuscript regardless of its state of publication in a scientific journal, and thus including also working papers, master's or PhD's theses, policy reports, etc.). The list of available models was recovered both exploiting common research databases as well as via an online survey administered to a number of Italian economists, who provided valuable information on the models they developed as well as indicated us additional models to be included in the review.

freely accessible; if the model converts income information from net to gross income; and, finally, the main bibliographic reference.

Table 1 – National static microsimulation models in Italy

Model	Creation	Main source of income information	Updates	Last update	In operation	Financing	Access on request	From net to gross	Bibliographic reference
ITAXMOD	1988	SHIW	Biennial	2009	No	Self-financed	No	Yes	Di Biase et al (1995)
TBM	1988	SHIW	Biennial	1996	No	Self-financed	No	Yes	Rizzi (1990)
MASTRICT	1996	SHIW	Biennial	2010	No	Self-financed	No	Yes	Proto (1999)
EUROMOD-IT	1997	SILC	Annual	2015	Yes	EU Commission DG-EMPL	Yes	No	Sutherland and Figari (2013)
MAPP	1998	SILC	Biennial	2015	Yes	Self-financed	No	No	Baldini et al. (2015)
TABEITA	1998	SILC	From 2007 biennial	2015	Yes	EU Commission DG-EMPL	No	Yes	Fiorio (2009)
ECONLAV	2004	SHIW	Annual	2014	No	MEF, MLPS and self-financed (ISFOL)	No	Yes	Cipollone et al., (2013)
MOD. PELLEGRINO	2000	SHIW	Biennial	2014	Yes	Self-financed	No	Yes	Pellegrino et al. 2011
SM2	2004	SILC and RA	Annual	2015	Yes	Self-financed	No	Yes	Betti et al. (2010)
MICROREG	2008	SILC	Annual	2015	Yes	Tuscany region	No	Yes	Maitino et al., (2008)
ITAXSIM	2011	SILC, INPS, RA	Annual	2015	Yes	Self-financed	No	Yes	www.dt.tesoro.it/[1]
MOD. FINANCE DEPT.	2012	SILC, RA, cadastral data	Biennial	2016	Yes	Self-financed	No	No	Di Nicola et al. (2015)
FAMIMOD	2014	SILC	Biennial	2014	Yes	Self-financed	No	No	Cozzolino et al. (2015)
BETAMOD	2015	SILC	-	2015	Yes	Self-financed and Fondazione Farma-factoring	No	Yes	Albarea et al. (2015)
MOD. UPB	2016	SHIW	Biennial	-	Yes	Self-financed	No	Yes	Gastaldi et al. (2017)

Note: (1) There is not a scientific publication of the ITAXSIM model, but some information is available on the website specified.

The first models implemented in Italy date back to the late Eighties and the Nineties: e.g., TBM by the University of Pavia (Rizzi, 1990) and ITAX-MOD by the "*Istituto di studi e analisi economica*" (ISAE) (Di Biase et al., 1995) were developed in 1988. Only few models of this first 'wave' of Italian

microsimulation models are still in operation (i.e., EUROMOD-IT, MAPP, TABEITA).

The vast majority of the listed models are based on households' survey data, namely on the Statistics on Income and Living Conditions (SILC) or the Survey of Italian Household Income and Wealth (SHIW) surveys. These surveys collect data on declared individual incomes. If is well-known that the quality of microsimulation models is related to the availability of good quality data (Sutherland, 1991), it is equally well-known that survey data and related declared incomes do have some limits (Di Nicola et al., 2015). First, survey data are plagued by problems of under-reporting, over-reporting and misclassification. Second, survey respondents are asked to report net incomes and therefore models have to apply algorithms to simulate the gross income, this possibly introducing further noise in the data. Third, income information in survey data is not as detailed as information collected in tax return forms (Di Nicola et al., 2015).

To improve the quality of data, some of the models complement survey data using other sources. For example, MicroReg corrects the incomes coming from SILC survey using data on income tax return, cadastral data as well as other survey data on the Italian households' consumption survey (Maitino et al., 2013). The main issue arising when joining two or more distinct data sources is the impossibility in achieving a perfect match of individuals and/or households in the different databases due to the unavailability of a unique identifier. Overcoming this problem is not easy and viable solutions are necessarily cumbersome and sub-optimal, namely use aggregate data to correct the original declared data or implement complex matching procedures to attach income values obtained from one dataset to the most similar individuals interviewed in the second sample based on some observable characteristics. Table 1 indicates the models for which income datasets are successfully linked with exact matching[2].

Among the models that are still in operation, most of them are updated every year or every two years. This means that a lot of resources are invested to create national models – usually self-financed – but they are not accessible to different users. In this regard, EUROMOD-IT is an exception, as it is freely accessible upon request.

Since Italian regions are competent in fiscal and social policy matters, tax-benefit models tailored to the region specificities could be used to carry out regional ex-ante evaluations. In this spirit, TREMOD was created. Before

[2] Examples of exact matching between SILC data and tax returns forms are the model of the Department of Finance, SM2 and ITAXSIM.

describing the features of this model, let us first briefly describe the state of the art of other regional models. Models developed at the regional level are often based on a national model using the sample related to the region of reference. An example is MicroReg (Maitino et al., 2013) or Puglia's model (Brunori and Lagravinese, 2012) which, using SILC data, allow for an estimate of policies at the regional level. The main threat these regional models have to face is the small size and the non-representativeness of the sample.[3] To partially redress this limitation, in a study carried out in Puglia and Emilia Romagna using SHIW data, Morciano (2006) pools together three surveys. Clearly, this solution increases sample sizes but the gains in terms of representativeness of the sample are not clear. Hence, the non-representativeness of regional sub-sample derived from national surveys is a serious threat to regional models. The availability of local surveys offers a way out allowing to obtain local microsimulation models that are based on representative samples of the population of interest. Unfortunately, local surveys that are suitable to the purposes of a microsimulation model are quite rare. An exception is the model developed by Baldini et al. (2004) for the municipality of Modena, based on the microsimulation model MAPP linked with the local survey on households' life conditions in Modena (ICESmo).

3. TREMOD

3.1. The model

The microsimulation model for the Province of Trento (TREMOD) is based on the EUROMOD platform and the Italian policy systems simulated in EUROMOD. EUROMOD is an open source tax-benefit microsimulation model developed since 1998 as part of the European programme, Targeted Socio-Economic Research, and is currently funded by DG Employment, Social Affairs and Inclusion. The project is led by an academic consortium representing several European countries coordinated by the Institute for Social and Economic Research (ISER). The EUROMOD project offers a set of algorithms able to reproduce tax and benefit policy rules for any of the EU-28 countries. Specifically, the EUROMOD modules allow one to simulates individual and household tax liabilities and benefit entitlements taking into account the policy rules in place in each member state (see also Sutherland and

[3] In section 5, we show that in the province of Trento the sample of SILC is not as representative as the local survey ICFT.

Figari, 2013). For example, EUROMOD allows to simulate income taxes (national and local), social contributions (paid by the employees, self-employed and employers), family benefits, housing benefits, social assistance and other income-related benefits. It is regularly updated. The information contained in the standard platform of EUROMOD derives from surveys (for example, EU-SILC), combined with information simulated by the model itself (for example, taxes and subsidies). EUROMOD can be used to answer a wide range of research questions. To date, it has been used to build microsimulation models in different national contexts: for example, South Africa (Wilkinson, 2009), Russian Federation (Popova, 2012) and the Republic of Serbia (Zarković-Rakić, 2010). To our knowledge, EUROMOD has rarely been adapted to sub-national contexts (Decancq et al., 2012). The definition and construction of the model that we propose, TREMOD, is based upon the Italian component of EUROMOD (Ceriani et al., 2013). This ensures a detailed reconstruction of the national legislation with regard to income taxes, assets and social security contributions, taking into account regional variations (e.g., additional regional income tax). Moreover, in EUROMOD are simulated social transfers, such as family and social allowances.

3.2. The data sources of TREMOD

The starting point for the construction of our microsimulation model is the availability of micro data on individuals' working conditions and social security. This first version of TREMOD uses 2010 data from two different sources: *Indagine sulle condizioni di vita delle famiglie trentine* (ICFT data) and administrative data on individual income tax returns provided by the Italian Revenue Agency (RA). The main source of data used for the construction of TREMOD is the *Indagine sulle condizioni di vita delle famiglie trentine* (ICFT). The ICFT is a survey containing rich socio-economic information at the individual and household levels. ICFT is a recall survey based on interviews with subjects older than 18 years old and resident in the province of Trento. In this first version of TREMOD, data are limited to the sixth wave of the survey carried out in 2010 on 7,200 respondents. The second source is individual income tax returns 2011 (*Dichiarazioni dei Redditi 2011, Modello Unico, 730 and 770*), thus containing the individual data for year 2010. It is worth noting that gross income is underestimated for those individuals who only own their 'main residence' or have property income lower than 500 euro on top of employment income and submit only a 770 form, as property income is not included in 770 return forms.

The exact matching of the two databases has been possible by using individual tax codes as a matching key. The entire process of building the basic input dataset for TREMOD was supervised by the Statistical Office of the Province of Trento, which ensured the compliance of all activities with data protection and privacy legislation throughout the entire process.

The TREMOD database is divided into three sections. The first section includes socio-demographic variables, with information on individual-level characteristics such as age, gender, level of education attained, and at the family level, the composition and kinship relationships. The second section includes variables concerning the employment status of the individual and her working history in terms of months worked, months of retirement, months of unemployment, as declared at the time of the interview. The third section contains all income variables that allow the reconstruction of total gross income, such as income from employment, retirement income, income derived from property and land, income from investments, earnings from other self-employment, private transfers received (such as checks from ex-spouses, for example).[4]

3.3. The validation of the data source

The use of survey data is critical when they are employed to draw inferences at a more general level (i.e. population). In this case, it becomes crucial to perform 'validation' checks on the survey data in order to verify their representativeness of the total population. To accomplish this, we compared the distributions of some socio-economic, labour market and income variables available in the sample that we use in TREMOD (ICFT data) with data from external official sources (Statistical Office of the Province of Trento, National Institute of Statistics ISTAT, and Italian Revenue Agency). The latter served as a reference (since official data sources are based on larger samples and are thus characterized by less variability, or they refer to the entire population) to validate the sample that we utilized. If large discrepancies emerged from this comparison, then problems in generalizing the results obtained using survey data would materialize. This would have negative implications on the analysis performed using the ICFT survey and, consequently, on the credibility of the resulting policy indications deriving from TREMOD.

[4] All the variables of the first two sections are derived from *Indagine sulle condizioni di vita delle famiglie trentine*, while those of the third section came from the administrative records of individual income tax returns.

Moreover, it is particularly important to verify the representativeness of the sample data with respect to different socio-demographic segments of the population (particularly by age class, or by employment status), since these may be recipients of specific public policies.

Among the socio-demographic characteristics available in the ICFT data comparable with external data sources, we had: age, gender, nationality, household composition, education level and employment status of individuals. The estimates of the distributions of these variables were performed using the weight provided by the Statistical Office of the Province of Trento (henceforth 'original weight')[5].

Table 2 – Number of residents in the Province of Trento by age classes. Comparison between ICFT data and External data

Age	Freq. ICFT data (1)	Freq. External data (2)	Ratio % (1/2)
Aged less than 15	73,300	80,469	91.1
15-24	58,693	52,131	112.6
25-34	51,912	64,328	80.7
35-44	74,440	85,304	87.3
45-54	79,203	77,682	102.0
55-64	75,183	63,497	118.4
More than 64	106,802	101,415	105.3
Total Population	519,533	524,826	99.0

Note: FBK-IRVAPP calculations. Data refer to the province of Trento in year 2010. The original weight is used for ICFT data. The external data source is ISTAT (www.demoistat.it).

Table 2 compares estimates on the number of residents in the Province of Trento, grouped by age classes, from the ICFT data and from the external source. The first two columns of the table show the frequencies of the individuals, for different age groups, as found in the survey and in the external source respectively. The third column is the ratio between the two different sources (multiplied by 100). A value of 100 for the ratio would show perfect comparability in the number of persons between the ICFT data and the external data source; lower values of the ratio would indicate a problem of under-representation of the sample; vice versa, values greater than 100, would indicate the existence of an over-representation.

Table 2 shows that the main differences between the survey sample and official sources, as regards age classes, consist in the under-representation of

[5] The Statistical Office of the Province of Trento built the weights based on the following variables: size of households, residence, presence of foreign citizens and people aged over 65 in the household.

young people aged 25-34 years (of whom there are 51,912 in the survey sample instead of 64,328 in the external data source) and in the over-representation of individuals aged 55-64 years (75,183 individuals in the ICFT instead of 63,497 in the external source). The percentage differences between the ICFT data and the external source, in these cases, are respectively equal to -19.3% and +18.4%. As regards the other age groups, problems are more limited. The number of persons in the 45-54 age class, and the over-65s, is very well represented, while the estimate is less precise, though overall acceptable, for the remaining bands (with differentials in representation close to 10%)[6].

Table 3 – Distribution of socio-demographic characteristics. Comparison between ICFT data and External data

Socio-demographic characteristics	Freq. % ICFT data (1)	Freq. External data (2)	Ratio % (1/2)
Non-Italian citizen	8.9	9.2	97.2
Gender (women)	52.2	51.1	102.2
Single-person households	30.1	32.9	91.4
Couples with children	38.7	36.8	105.2
Couples without children	23.5	24.9	94.6
Single-parent households	7.7	5.4	142.2
Primary education or no qualification	17.6	18.0	97.9
Lower secondary education	29.0	28.7	101.1
Upper secondary education	41.0	41.5	99.0
Tertiary degree or above	12.3	11.8	104.1

Note: FBK-IRVAPP calculations. Data refer to the province of Trento in year 2010. The original weight was used for ICFT data. The external data source is ISTAT, for survey data the provided weights are used.

Other socio-demographic characteristics that could be used to validate the ICFT survey were: citizenship, gender, household composition and the level of education of individuals. Table 3 shows that, for most of the cases, the ICFT source provides highly representative data of the population. The ICFT is particularly able to represent the composition of the provincial population of Trento by citizenship: the estimate of the proportion of individuals with non-Italian citizenship is 8.9% compared to 9.2% resulting from the external

[6] The problem of representation of individuals by class of age in survey samples comes as no surprise here, given that participation in the survey is voluntary and the response rate often differs by age levels. Generally, to avoid problems of distortion in the sample, the problem is addressed at a final stage by the application of appropriate techniques of post-stratification (this is what we propose later, see section 3.3).

source. The ICFT data are also highly representative in terms of gender composition: the proportion of women in the sample is 52.2%, only slightly higher than the 51.1% resulting from external statistics.

Regarding the variables measuring household composition, survey data are representative, in large part, of the number of couples with children (38.7% of total households instead of 36.8%), couples without children (23.5% instead of 24.9%) and the number of people who live alone or who do not form a family (single-person households are 30.1% instead of 32.9%). A problem arises, however, in regard to the representativeness of single-parent households, which are over-estimated in the ICFT survey (7.7% instead of 5.4% of households), though with a large variability.

Very convincing results concern the ability of the ICFT survey to reproduce the distribution of educational qualifications across the population. The representativeness of the number of individuals who had received no education, only primary level education, lower or upper secondary education and tertiary education has no critical issues, in that frequencies in the ICFT data are very similar to those of ISTAT, apart from a slight over-estimation of persons with tertiary qualifications.

Table 4 – Labour-market indicators. Comparison between ICFT data and External data

Labour-market participation	Freq. % ICFT data (1)	Freq. External data (2)	Ratio % (1/2)
Activity rate	62.6	69.0	90.7
Employment rate	58.8	66.0	89.1
Unemployment rate	5.9	4.3	138.0
Unemployment rate (15-24)	21.4	15.1	141.6
Activity rate (Women)	64.3	60.5	106.2
Employment rate (Women)	59.6	57.3	104.0
Unemployment rate (Women)	7.3	5.2	140.2

Note: FBK-IRVAPP calculations. Data refer to the province of Trento in year 2010. The original weight was used for ICFT data. The external data source is ISTAT (Labour Force Survey) and the relative sample weights are used.

A further group of variables used to verify the representativeness of the ICFT survey with respect to the values from external sources were those concerning labour market characteristics: in particular, the labour market status of the respondents. Table 4 shows the employment, activity and unemployment rates (defined for the total population, young people aged 15-24 years old and women respectively) in the ICFT (column 1) and in the external ISTAT source (column 2). Comparison between the two data sources shows that the ICFT

data under-represents active persons (62.6% instead of 69.0%) and the employed (58.8% in the ICFT compared to 66.0% from ISTAT). On considering the same indicators by gender, also evident is a slight over-representation in the number of women interviewees who belonged to the labour force.

The unemployment rate also appears to be over-estimated. In the ICFT the unemployment rate among respondent is 5.9%, while the value from the external source is 4.3%. Women are unemployed in 7.3% of cases in the ICFT survey versus 5.2% in the ISTAT data and, similarly, the rate of youth unemployment in the ICFT is 21.4% against 15.1% from ISTAT[7].

Table 5 – Number of income recipients and total income (in thousands of euros) referred to main sources of income. Comparison between ICFT data and Population data

Source of income	Freq. ICFT data (1)	Total income ICFT data (2)	Freq. Population (3)	Total income Population (4)	Ratio % (1/3)	Ratio % (2/4)
Labour income	206,234	4,164,646	239,591	4,426,109	86.1	94.1
Self-employment income	11,321	201,996	14,022	323,119	80.7	62.5
Retirement income	142,255	2,309,307	135,183	2,036,244	105.2	113.4
Disposable income	373,556	6,668,324	408,046	6,858,059	91.5	97.2
Incomes from property	206,149	345,534	200,467	354,766	102.8	97.4

Note: FBK-IRVAPP calculations. Data refer to the Province of Trento in year 2010. The original weight was used for ICFT data. Data on income provided by the Italian Revenue Agency.

Finally, we controlled for a set of variables of interest concerning income by considering four categories of interest: labour income (for employee), self-employment income, retirement income (pensions) and disposable income (i.e. income after tax payments and with non-taxable transfers such as family allowances and social pensions). As described above (Section 3.1), this information was obtained from the Italian Revenue Agency (RA) providing income data for the year 2010. Table 5 shows the measures of income that we considered in two different ways. The first was obtained by linking information on income from the Italian Revenue Agency to individuals interviewed in the ICFT survey. For this case, column 1 shows the number of income recipients based on different types of income, and the sum of the income that they declared (referred to as 'total' in column 2). The second

[7] The increased frequency of unemployed people interviewed in the ICFT survey may be a consequence of their easier availability and willingness to respond to the interview.

way concerned income measures for the total population from official administrative archives. Columns 3 and 4 indicate the same information as from the ICFT survey (columns 1 and 2) for external sources, which, here, is relative to the entire population and not limited to the subjects interviewed by the ICFT.

Comparison of the number of subjects declaring different sources of income in the ICFT and in total population (penultimate column) shows an underestimation by the ICFT survey of persons reporting labour income (about 14% less than in the population) and self-employment income (about 20% less). Also partially over-estimated is the representativeness of the number of pensioners in the ICFT survey, which are present in excess (+5.2%) with respect to the population value.

Considering the total amount of income that respondents declared, the last column in Table 5 shows the ratios between the amounts declared by persons in the ICFT and in the total population. Labour incomes are under-estimated by 5.9%. By contrast, incomes from pensions are over-estimated by 13.4%. The most critical fact, however, concerns the amount of incomes related to self-employed persons, which, according to the ICFT survey, turn out to be much lower than those of the population (about 40%). In other words, the ICFT survey, besides under-representing the large number of self-employed workers, also largely under-estimates their income. A comforting result comes, however, from the disposable income of the respondents: this turns out to be representative, with a very small deviation (-2.8%) from the entire population value.

3.4. The post-stratification procedure

As described in the previous section, the ICFT survey is unable to represent correctly population totals with respect to some demographic, social and income aspects of importance when performing microsimulation analysis. The problem is evident with regard to population by age classes, employment status and number and incomes of self-employed workers. Other differences, though less significant, regard household composition and labour market indicators of respondents to the survey. Hence, we correct the sample in the ICFT via re-balancing for such characteristics using a 'post-stratification' statistical procedure. By means of this procedure, it is possible to build appropriate weights readjusting the sample size in order to make them more representative of the population data (Judkins, 1990).

The starting point of the correction process was the re-proportioning of the original weight released by the Statistical Office of the Province of Trento. Unfortunately, this type of procedure is not without problems. Indeed, controlling for balancing along one dimension (variable) may lead to a correction of another one in the opposite direction. In other words, the correction along one or more marginal distributions does not guarantee the correctness of the joint distribution. Moreover, the problem is larger, the greater the number of variables considered in the post-stratification algorithm.

Table 6 shows the percentage ratio between the sample size in the ICFT and from external sources according to socio-demographic variables, household composition and status in the labour market, as evidenced by the use of different weights for post-stratification. In particular, four different weights were calculated which, incrementally, increased the number of variables inserted in the post-stratification procedure. The first weight (Weight 1) took into account the main socio-demographic variables (age, nationality and gender of individuals). Weight 2 included the level of education to the control of the characteristics just mentioned. Weight 3 added to Weight 2 the characteristics of individuals related to their labour-market status. Finally, Weight 4 added the representativeness for the number of persons self-employed.

As shown in Table 6, the post-stratification procedure allowed important improvement to be made to the quality of the data available. The increase in the correspondence between the ICFT and the external source proved to be constant with respect to all the weights specifications and, despite the inclusion of numerous variables, the distributions of the different age classes in the ICFT data and in the population were fully coincident.

Table 6 shows that, also with regard to gender and citizenship, the differences between the observed frequencies in the ICFT data and those found in external sources were completely corrected.

Household characteristics were not included directly in the post-stratification procedure due to the limited degrees of freedom (impossibility to control for any variable in the optimization process). However, following the application of different weights, differences in the distribution of such variables with those from external sources differences were very limited. More effective was the over-representation of single-parent households, as already found by the application of the original weight. As previously mentioned, this error was, however, not crucial given the limited size of the phenomena analyzed.

Table 6 – Distribution of socio-demographic characteristics and labour-market participation indicators. Percentage ratios between ICFT data (different weights) and External data

Variables	Original Weight	Weight 1	Weight 2	Weight 3	Weight 4
Aged less than 15	91.1	100.0	100.0	100.0	100.0
15-24	112.6	100.0	100.0	100.0	100.0
25-34	80.7	100.0	100.0	100.0	100.0
35-44	87.3	100.0	100.0	100.0	100.0
45-54	102.0	100.0	100.0	100.0	100.0
55-64	118.4	100.0	100.0	100.0	100.0
More than 64	105.3	100.0	100.0	100.0	100.0
Total Population	99.0	100.0	100.0	100.0	100.0
Nationality (not Italian)	97.2	100.0	100.0	100.0	100.0
Gender (women)	102.2	100.0	100.0	100.0	100.0
Single-person household	91.4	90.0	90.0	90.3	90.5
Couples with children	105.2	108.5	108.4	108.3	108.1
Couples without children	94.6	91.3	91.3	90.9	90.9
Single-parent households	142.2	143.3	143.7	144.4	144.2
Primary education or no qualification	97.9	91.5	100.0	100.0	100.0
Lower-secondary education	101.1	97.7	100.0	100.0	100.0
Upper-secondary education	99.0	102.3	100.0	100.0	100.0
Tertiary degree	104.1	110.6	100.0	100.0	100.0
Activity rate	90.7	94.9	94.4	97.4	97.4
Employment rate	89.1	93.3	92.7	97.3	97.3
Unemployment rate	138.0	136.2	138.9	101.5	101.1
Unemployment rate (15-24)	141.6	140.5	141.0	106.2	105.1
Activity rate (Women)	106.2	109.7	108.4	111.6	111.5
Employment rate (Women)	104.0	107.4	106.1	111.5	111.4
Unemployment rate (Women)	140.2	139.1	141.6	104.2	103.6

Note: FBK-IRVAPP calculations. The data refer to the province of Trento for 2010. The first weight (Weight 1) takes into account age, nationality and gender. Weight 2 sums the education level to the control of the characteristics just mentioned. Weight 3 adds to Weight 2 characteristics of individuals related to their labour-market status. Finally, Weight 4 adds the representativeness for the number of self-employed individuals (data sources are presented in Section 3.2).

The education characteristics of the respondents were – as in the case of variables related to age, gender and nationality – fully representative of the actual values after the use of Weight 2, which took account of the education variables in the calculation process.

The post-stratification procedure affected the labour-market variables starting from the application of Weight 3, and it proved effective in correcting the initial imbalance between ICFT data frequencies and the external

source. The aggregate unemployment rate (but also disaggregated by groups for young people and women) was reduced to percentage values close to those of the external source. We thus overcome a problem that, while analyzing small groups (unemployed people is a limited share of the total population), may be particularly relevant to the analysis of policies in Trentino, since many public policies specifically target the unemployed. Employment and activity rates largely corresponded to the reference data, with a slight over-estimates only in respect of the indicators related to gender.

Table 7 – Ratios between ICFT data (different weights) and overall population data on individual income tax returns (different weights). Income earners number and total amount of main sources of income

Source of income	Original weight	Weight 1	Weight 2	Weight 3	Weight 4
A. Ratio between frequencies					
Labour income	86.1	90.0	89.3	91.1	90.8
Self-employment income	80.7	82.4	79.2	77.7	99.5
Retirement income	105.2	97.0	97.6	96.6	96.5
Disposable income	91.5	91.7	91.6	92.3	92.3
B. Ratio between total amount					
Labour income	94.1	98.5	96.7	99.9	99.3
Self-employment income	62.5	64.0	60.0	60.8	78.0
Retirement income	113.4	104.7	102.7	100.9	100.8
Disposable income	97.2	98.0	96.5	98.1	98.4

Note: FBK-IRVAPP calculations. Data refer to the province of Trento for 2010. The first weight (Weight 1) takes into account age, nationality and gender. Weight 2 sums the education level to the control of characteristics just mentioned. Weight 3 adds to Weight 2 characteristics of individuals related to their labour-market status. Finally, Weight 4 adds the representativeness for the number of self-employed individuals (data source is presented in Section 3.2).

Table 7 shows, respectively, the relative percentages of the number of individuals (Part A of the table) and the total amounts of income (Part B of the table) by type of income declared by individuals. Weight 4 corrected for the distortion due to the under-representation of the number and the average income level of self-employed workers, including the balance condition for the number of self-employed workers in the optimizing process. The application of Weight 4 significantly increased the number of self-employed workers represented in the ICFT survey. The large number of individuals reporting other types of income (employees, retirees and disposable income) remained relatively closed to the actual population frequency. Despite the correction for the

relative frequencies, Table 7 shows an incremental improvement in the ability to represent incomes of self-employed people even if it cannot bring their representativeness to entirely satisfactory levels. Income from self-employment is thus under-estimated compared to the values for the population, while those of employee income and retirement income, as well as disposable income, are instead good approximations of the actual values.

Overall, the post-stratification procedure described made it possible, through Weight 4, to rebalance the data source in the ICFT data in order to make it particularly suited to representing the population of the Province of Trento on most of the variables considered essential for policy analysis using TREMOD: in particular, the representation of different segments of the population (age classes), the distribution by gender and nationality, level of education, household composition and employment status. The number and the income situation in the ICFT sample, moreover, are now able to effectively capture the actual conditions of the population of the Province of Trento particularly in regard to employees and pensioners. There remain some critical issues relative to the amount of declared income by self-employed workers, which is partly under-estimated. Unfortunately, numerous dimensions were involved in the post-stratification process, and the results clearly cannot account for everything at the same time. However, the starting conditions were significantly improved through the stratification process.

4. Assessing TREMOD's performance

TREMOD makes it possible to simulate the amount of income taxes and social transfers and to calculate individual or household disposable income starting from information on total income and family composition. The performance analysis of estimates produced using TREMOD, described in this section, sought to verify that each of the components calculated by the microsimulation model (for example taxes, tax credit and deductions) was equivalent to real data. More precisely, our analysis focused on the ability of our model to calculate total income, taxable profit, gross and net taxes, and various types of tax credits and deductions[8].

[8] Social transfers were not available in administrative data for 2010, therefore it is not possible to compare the simulated benefits with the real data. Property taxation was imputed from RA data and not simulated by the model. Municipal surtaxes were not included as the data for 2010 were not available. Such information will be included in the 2012 update of the model.

Table 8 – Analysis of simulated fiscal variables using TREMOD. Comparison of various sources of income between TREMOD estimates and administrative data (RA) based on persons in the ICFT survey. Income earners number (Freq.) and average amount (in Euro)

Source of income	Freq. TREMOD (1)	Mean TREMOD (2)	Freq. RA (3)	Mean RA (4)	Ratio % (1/3)	Ratio % (2/4)
Total income	380,216	21,285	380,216	21,268	100.0	100.1
Other deductions	170,127	1,610	169,770	1,613	100.2	99.8
Net taxable income	379,084	21,055	376,094	20,561	100.8	102.4
Gross Tax	376,672	5,482	367,884	5,618	102.4	97.6
Net tax	302,679	4,824	307,715	4,785	98.4	100.8
Tax credit for income source[1]	355,465	1,127	343,825	1,138	103.4	99.0
Family tax credits[1]	92,725	1,075	91,173	1,074	101.7	100.1
Tax credit (dependent spouse/children) [1]	92,269	1,069	90,910	1,064	101.5	100.5
Other tax credits	238,718	632	238,124	631	100.2	100.2
Disposable income	379,084	17,292	380,216	17,396	99.7	99.4

Note: FBK-IRVAPP calculations. Data refer to the Province of Trento for 2010. Weight 4 was used to correct data freq. (see Section 3.3). Data on income provided by the Italian Revenue Agency. (1) For these tax credits the tax unit is the household instead of the individual.

We first compared the estimates obtained with TREMOD with the real values gathered from individual income tax returns for individuals present in the ICFT survey. Table 8 shows that TREMOD's estimates were very similar to the real values along all the dimensions that we took into account. The relationship between TREMOD's estimates and those of the individual income tax returns are always very close to 100. This result demonstrates the capacity of TREMOD to accurately estimate individual tax and social security benefits.

Furthermore, since the purpose of TREMOD is to provide estimates valid for the entire population, it should be verified whether the results of TREMOD are close to the tax returns data of the entire population (Table 9). To compare survey and population data, we used the 'optimal' weight (i.e., Weight 4), since it included all relevant dimensions, as seen in Section 3.3. Table 9 shows that the number of people who declare a total income (of any type) is slightly underestimated, 92.3% compared to the actual population, while the amount of total income is close to 100%. This error influences the relationship between the frequencies, but it is only marginally relevant when

we consider income amount instead of the number of individuals. TREMOD also performs well in simulating other variables (i.e., taxable income, gross tax, net deductions from work and retirement, disposable income), albeit with slight differences. Table 10 shows that TREMOD performs remarkably well in simulating a number of inequality indices, thus proving to be a reliable instrument to assess the redistributive effects of public policy.

Table 9 – Analysis of simulated fiscal variables using TREMOD. Comparison of various sources of income between TREMOD estimates and administrative data (RA) referring to total population. Income earners number (Freq.) and declared total amount (in thousands of Euro)

Source of income	Freq. TREMOD (1)	Total TREMOD (2)	Freq. RA (3)	Total RA (4)	Ratio % (1/3)	Ratio % (2/4)
Total income	380,505	8,100,595	412,367	8,218,799	92.3	98.6
Taxable income	379,373	7,989,332	407,555	7,873,687	93.1	101.5
Gross tax	376,961	2,066,770	399,786	2,110,699	94.3	97.9
Net tax	302,892	1,461,191	326,563	1,515,161	92.8	96.4
Tax credit for income source	355,729	400,676	370,366	421,017	96.0	95.2
Disposable income	379,373	6,561,434	412,367	6,703,856	92.0	97.9

Note: FBK-IRVAPP calculations. Data refer to the province of Trento for 2010. For TREMOD data we used weight 4 as population size weight (see Section 3.3). Data on income provided by the Italian Revenue Agency.

Table 10 – Selected inequality indices simulated by TREMOD and estimated on administrative data (RA)) referring to total population

Index	TREMOD	RA
Pre-tax Gini	0.41	0.42
Post-tax Gini	0.37	0.37
Average tax rate	0.19	0.19
Reynolds-Smolensky net redis. effect	0.04	0.04
Kakwani progressivity index	0.18	0.19

Note: FBK-IRVAPP calculations. The TREMOD simulations were obtained using 'weight 4' (see Section 3.3). Income data comes from the Italian Revenue Agency archive.

5. A comparison between TREMOD and EUROMOD

The most innovative aspect of TREMOD compared to the Italian component of EUROMOD consists in the data sources that we used. TREMOD's database is about ten times larger than the sub-sample of residents in the Province of Trento in the IT-SILC survey (the data source used in EUROMOD), and it is matched with administrative data that provide more detailed and precise information than self-reported declarations. The matching of survey and administrative data has a huge potential in terms of future developments of the input database and the model that will be discussed in the conclusions (see Section 6).

To assess the specific gains in using TREMOD instead of EUROMOD, we compared results obtained with the two models. Table 11 compares the estimates of the number of income recipients and the average amount of income obtained using TREMOD and EUROMOD using SILC data (subsample SILC-TN for the Province of Trento) and the Italian Revenue Agency (RA), for different segments of the total population (some relevant age classes, and gender).

On average, TREMOD estimates of total and disposable income show lower deviations from official statistics data of the Italian Revenue Agency (RA) than do SILC-TN estimates, even though the net tax estimated in SILC-TN is closer to the RA. The standard errors show that the SILC-TN estimates are more variable, partly because of the smaller SILC-TN sample size where 793 respondents are present in the survey, while the ICFT involved 7,200 in 2010. Overall, the results obtained from the two samples are quite satisfactory, even if real population values sometimes do not fall within the confidence intervals obtained from the estimates performed using TREMOD and SILC-TN samples.

The Tables 12 and 13 show comparisons between estimates using TREMOD, EUROMOD (subsample SILC-TN for the Province of Trento) and the Italian Revenue Agency (RA) by percentiles of the income distribution.

Table 11 – Comparison among the results of TREMOD, EUROMOD-IT (subsample Province of Trento SILC-TN) and Italian Revenue Agency (RA) data. Number of income earners (Freq.), average amount (in euros)

Source of income	Freq. TREMOD	Mean TREMOD	Freq. SILC-TN	Mean SILC-TN	Freq. RA	Mean RA
A. Population						
Total Income	380,505	21,289	400,425	21,845	412,367	19,931
		(290)		(731)		
Disposable Income	379,373	17,295	402,686	17,710	412,367	16,257
		(197)		(539)		
Net tax	302,892	4,824	331,010	4,519	326,563	4,640
		(120)		(239)		
B. Age class 25-45						
Total Income	132,429	21,500	139,731	23,605	143,355	19,569
		(391)		(967)		
Disposable Income	132,127	17,744	139,731	19,267	143,355	16,181
		(277)		(728)		
Net tax	113,031	4,276	119,976	4,499	117,672	4,129
		(137)		(306)		
C. Age class 46-65						
Total Income	125,836	26,379	129,558	27,400	134,800	25,444
		(615)		(1,615)		
Disposable Income	125,453	20,943	131,819	21,586	134,800	20,242
		(400)		(1,168)		
Net tax	106,833	6,169	113,742	5,806	113,766	6,164
		(257)		(530)		
D. Men						
Total Income	193,121	26,585	199,697	27,641	213,289	24,196
		(493)		(1,178)		
Disposable Income	192,947	21,114	199,697	22,274	213,289	19,302
		(325)		(849)		
Net tax	167,834	6,033	174,750	5,707	180,564	5,781
		(198)		(390)		
E. Women						
Total Income	187,384	15,831	200,728	16,079	198,966	15,362
		(257)		(728)		
Disposable Income	186,426	13,343	202,989	13,220	198,966	12,994
		(190)		(560)		
Net tax	135,058	3,322	156,260	3,191	145,894	3,229
		(93)		(227)		

Note: FBK-IRVAPP calculations. Standard errors in parentheses. Data refer to the province of Trento for 2010. For TREMOD data we used weight 4 as population size weight (see Section 3.3). SILC-TN are estimates based on the EUROMOD model (using as data source the subsample of individuals, in the IT-SILC survey, resident in the Province of Trento). Italian Revenue Agency (RA) data refer to the whole population of data on individual income tax returns.

Table 12 – Comparison among the results of TREMOD model, EUROMOD (subsample Province of Trento SILC-TN) and Italian Revenue Agency (RA) data. Percentiles 10, 50 and 90 (in euros)

Source of income	10% TREMOD	50% TREMOD	90% TREMOD	10% SILC-TN	50% SILC-TN	90% SILC-TN	10% NTI	50% NTI	90% NTI
A. Population									
Total Income	4,000	18,306	37,729	2,210	20,236	42,161	2,001	16,772	36,027
Disposable Income	4,099	15,664	29,240	1,415	16,721	32,930	1,915	14,627	28,445
Net tax	600	3,121	9,723	416	3,271	9,863	329	2,839	9,159
B. Age class 25-45									
Total Income	5,684	19,997	35,125	6,183	23,014	39,884	1,670	18,346	33,812
Disposable Income	5,589	16,908	28,209	5,810	18,961	31,194	1,578	15,863	27,172
Net tax	775	3,195	7,646	1,238	3,699	8,292	340	2,966	7,934
C. Age class 46-65									
Total Income	5,987	22,131	44,859	5,081	23,185	51,999	3,964	20,818	45,314
Disposable Income	5,986	18,654	34,314	2,443	18,753	41,775	3,771	17,812	34,681
Net tax	637	3,730	11,650	463	3,664	12,158	547	3,594	12,500
D. Men									
Total Income	6,754	22,224	45,983	7,475	24,325	50,003	2,093	20,126	43,498
Disposable Income	6,583	18,686	34,933	7,239	19,912	38,873	2,000	17,240	33,436
Net tax	952	3,672	12,565	746	3,846	11,501	528	3,407	11,656
E. Women									
Total Income	2,545	13,791	29,243	1,186	14,261	31,782	1,953	13,118	28,968
Disposable Income	2,639	12,193	23,445	653	12,282	25,403	1,854	11,848	23,532
Net tax	399	2,428	6,601	261	2,408	6,627	225	2,124	6,534

Note: FBK-IRVAPP calculations. Data refer to the province of Trento for 2010. For TREMOD data we used weight 4 as population size weight (see Section 3.3). SILC-TN are estimates based on the EUROMOD model (using as data source the subsample of individuals, in the IT-SILC survey, resident in the Province of Trento). Italian Revenue Agency (RA) data refer to the whole population of data on individual income tax returns.

Table 13 – Comparison among the results of TREMOD model, EUROMOD (subsample Trentino – SILC(TN)) and Italian Revenue Agency (RA) data. Percentiles 95, 99 and 99,9 (in Euro)

Source of income	95% TREMOD	99% TREMOD	99,99% TREMOD	95% SILC-TN	99% SILC-TN	99,99% SILC-TN	95% NTI	99% NTI	99,99% NTI
A. Population									
Total Income	49,813	100,210	472,828	51,047	87,050	217,918	49,202	103,705	534,857
Disposable Income	37,331	66,174	293,789	40,217	60,193	160,569	37,095	70,568	324,296
Net tax	14,757	35,343	179,039	13,055	27,431	57,349	14,695	37,461	231,011
B. Age class 25-45									
Total Income	46,426	87,134	130,396	44,431	62,905	128,550	44,587	82,734	485,502
Disposable Income	35,503	59,358	88,369	34,728	42,438	85,008	34,137	58,633	286,031
Net tax	12,735	29,102	42,348	11,501	18,532	43,542	11,991	27,488	203,735
C. Age class 46-65									
Total Income	61,947	136,149	472,828	76,824	109,211	217,918	66,498	132,706	643,790
Disposable Income	45,152	87,719	293,789	56,488	74,716	160,569	48,324	88,402	387,563
Net tax	17,822	48,175	179,039	20,379	37,127	57,349	21,130	48,198	276,924
D. Men									
Total Income	64,741	126,244	472,828	62,872	91,110	217,918	61,859	126,488	680,924
Disposable Income	46,354	84,962	293,789	45,889	62,748	160,569	45,318	84,501	404,000
Net tax	18,973	42,027	179,039	17,102	34,495	57,349	19,187	46,055	298,537
E. Women									
Total Income	35,760	61,947	162,281	35,904	49,136	97,086	35,540	66,484	302,308
Disposable Income	28,116	46,300	150,582	29,366	37,583	64,711	27,909	47,851	182,720
Net tax	9,159	21,141	56,730	9,774	14,747	35,920	9,302	22,748	121,001

Note: FBK-IRVAPP calculations. Data refer to the province of Trento for 2010. For TREMOD data we used weight 4 as population size weight (see Section 3.3). SILC-TN are estimates based on the EUROMOD (using as data source the subsample of individuals, in the IT-SILC survey, resident in the Province of Trento. Italian Revenue Agency (RA) data refer to the whole population of data on individual income tax returns

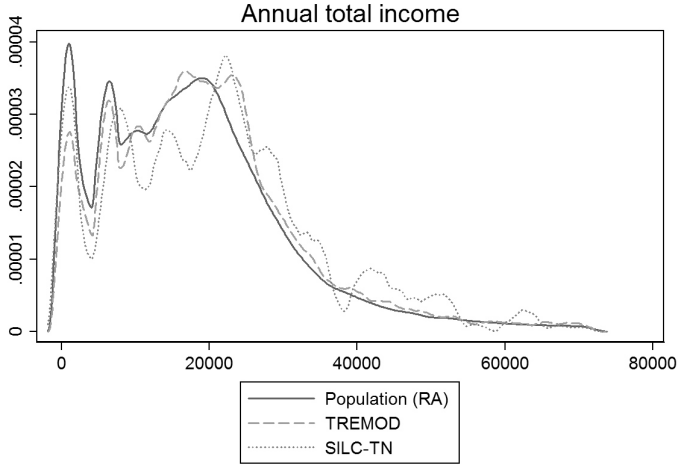
Table 12 shows that the lower part of the income distribution, and in particular the tenth percentile, is approximated better in SILC-TN than in TREMOD. This result, however, is less noticeable when stratifying the population by relevant age class and gender. Nevertheless, these results suggest caution in the analysis of fiscal policies targeted on individuals with lower incomes. Regarding the median and the ninetieth percentile, TREMOD's estimates better approximate the distribution of the population. The result is confirmed in subsamples obtained in relation to age and gender. This confirms that the ICFT survey is representative of the income distribution of the population of the Province of Trento with some limitations in the representativeness of the lower part of the income distribution.

As shown by Table 13, a significant advantage of TREMOD over EUROMOD (using SILC-TN) is that it is able to capture more accurately the distribution of the highest percentiles of the distribution. In recent years, the

economic literature has emphasized the importance of analysis of top incomes, focusing on the share of total income held by 5%, 1% or even 0.01% of the wealthiest population (see, for example, Atkinson et al., 2011). Given that the representativeness for the highest percentiles of the income distribution may be very limited in small samples, the TREMOD model, which is based on a dataset that contains over 7,000 interviews, may have some structural advantages compared to EUROMOD. Table 13 confirms that TREMOD is more accurate in representing the distribution of high incomes and therefore more suitable for analysing the impact of fiscal policies on the disposable income of the wealthiest segment of the population.

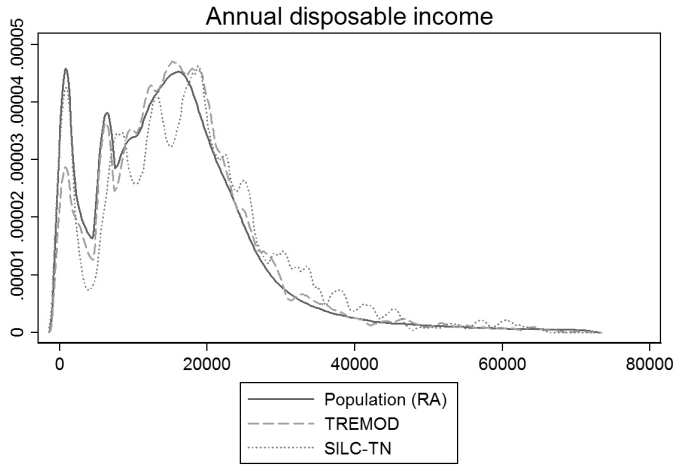
Thus far, we have analysed the quantiles and have produced some descriptive statistics of the income distribution. We now also provide a more detailed comparison based on the entire income distribution using non-parametric density estimates (see Silverman, 1986). Figures 1 and 2 below show the distribution of total and disposable income, respectively. The solid line represents the income density estimated on the population (RA data), the dashed line refers to TREMOD, and the dotted line is estimated on SILC-TN. Both figures show that the income distribution of the population is represented with greater precision using TREMOD rather than SILC-TN, except for individuals with low incomes.

Figure 1 – Total income: comparison among the results of TREMOD model, EUROMOD (subsample Province of Trento SILC-TN) and Italian Revenue Agency (RA) data



Note: FBK-IRVAPP calculations. Data refer to the province of Trento for 2010. For TREMOD data we used weight 4 as population size weight (see Section 3.3). SILC-TN are estimates based on the EUROMOD model (using as datasource the subsample of individuals, in the IT-SILC survey, resident in the Province of Trento). Italian Revenue Agency (RA) data refer to the whole population of data on individual income tax returns.

Figure 2 – Disposable income: comparison among results of TREMOD model, EUROMOD (subsample Province of Trento SILC-TN) and Italian Revenue Agency (RA) data



Note: FBK-IRVAPP calculations. Data refer to the province of Trento for 2010. For TREMOD data we used weight 4 as population size weight (see Section 3.3). SILC-TN are estimates based on the EUROMOD model (using as datasource the subsample of individuals, in the IT-SILC survey, resident in the Province of Trento). Italian Revenue Agency (RA) data refer to the whole population of data on individual income tax returns.

Finally, we focus on a synthetic analysis using the household income distribution. It should be noted that distribution statistics are not available from external sources and therefore it was not possible to validate the results obtained with TREMOD. We simply show some popular statistical distributions of equivalent household income by comparing the results obtained with each model. The disposable household income was obtained in order to take account of the different sizes of families, based on the so-called OECD-modified equivalence scale that assigns a weight equal to 1 to the householder, 0.5 to other adults (those older than 14) and 0.3 to minors.

Table 14 shows the Gini index and the individual and household poverty rate[9]. The two sample estimates of the levels of inequality and poverty are very similar.

[9] In accordance with a common convention (e.g see Eurostat, 2016), we considered as 'poor' those individuals living in households whose income did not reach 60% of the median equivalised disposable income of the reference population. The family poverty rate is the share of households below the poverty line, while the individual poverty rate is the share of individuals below the threshold.

Table 14 – Comparison between Gini index in TREMOD model and in EUROMOD- (subsample Province of Trento SILC-TN)

Sample	Gini index	Individual poverty rate (60% threshold)	Family poverty rate (60% threshold)
TREMOD	0,29	17,88	16,92
SILC-TN	0,27	17,45	17,80

Note: FBK-IRVAPP calculations. Data refer to the province of Trento for 2010. For TREMOD data we used weight 4 as population size weight (see Section 3.3). SILC-TN are estimates based on EUROMOD (using as data source the subsample of individuals, in the IT-SILC survey, resident in the Province of Trento).

6. Concluding Remarks

While microsimulation models at the national level in Italy are rich in number and methodologically sound, the panorama of microsimulation modelling at the regional level is far less advanced. In most cases, models developed at the regional level are simple adaptations of national models and are based on subsamples of national surveys. The main threats to these regional models are the small size and the non-representativeness of the samples. Hence regions, with only very few exceptions, are left with poor, if any, analytical tool to carry out ex-ante evaluations. Such a situation is particularly unfortunate as regions in Italy have not negligible competences regarding taxation and social policies. The paper aims at partially redressing this shortcoming by presenting a new regional microsimulation model and providing a detailed account of its development process.

More precisely, this paper illustrates the development and validation of TREMOD, the new "static" tax-benefit microsimulation model for the province of Trento model developed for the province of Trento. The model is based upon the EUROMOD platform and combines survey and administrative data of the resident population of the Province of Trento. In order to assess the success of this process and then evaluate the capacity of TREMOD to produce reliable estimates of policy relevant variables, various statistical analyses have been performed.

We have shown that TREMOD is robust with regard to some crucial aspects related to the reliability of microsimulations estimates. First, the database proved to be highly representative of the entire population of the Province of Trento with respect to many socio-economic and demographic indicators. Second, TREMOD proved to be highly reliable in the simulation of various quantities of interest, including different definitions of income, tax and tax credit.

Third, the analyses presented in this paper confirm the importance of using objective measures of income based on administrative sources (in our case tax returns) compared with self-reported information collected by surveys, since the latter are affected by larger measurement errors.

Concerning the potential applications of TREMOD, the model can be used to simulate the impact of a wide range of interventions in the fields of taxation and social policies at the local level. Such policies could include not only national policy interventions (e.g., unemployment benefits), but also the modifications introduced by the province of Trento to such national programmes (i.e., the province of Trento has introduced important modifications to the national legislation in these fields) as well possible variations in provincial income surtaxes. Moreover, as social policy legislation takes place also at the regional level, microsimulation models like TREMOD can be employed to support regional policy makers. The case of the province of Trento is illustrative on this regard, as the province manages a large number of social monetary transfers and benefits programmes, such as the *Reddito Minimo di Garanzia*. This programme has been operating since 2009 and is aimed at introducing a minimum guarantee income in Trentino. TREMOD can be employed to support decision making on such a programme provided that it will be fully integrated with the administrative archives containing key information on the programme implementation.

References

Albarea A., Bernasconi M., Di Novi C., Marenzi A., Rizzi D. and Zantomio F. (2015). Accounting for Tax Evasion Profiles and Tax Expenditures in Microsimulation Modelling. The BETAMOD Model for Personal Income Taxes in Italy. *The International Journal of Microsimulation*, 8(3): 99-136.

Atkinson A.B., Piketty T. and Saez E. (2011). Top Incomes in the Long Run of History. *Journal of Economic Literature*, 49(1): 3-71.

Baldini M., Bosi P. and Colombini S. (2004). Efficacia Selettiva dell'Ise nell'Erogazione di Prestazioni Sociali Agevolate nella Provincia di Modena. *CAPP Working Paper* 456.

Baldini M., Giarda E. and Olivieri A. (2015). A Tax-Benefit Microsimulation Model for Italy: A Partial Evaluation of Fiscal Consolidation in the Period 2011-2014. *Prometeia Nota di Lavoro*, 1.

Betti G., Donatiello G. and Verma V. (2010). The Siena Microsimulation Model (sm2) for Net-Gross Conversion of Eu-Silc Income Variables. *International Journal of Microsimulation*, 4(1): 35-53.

Brunori P. and Lagravinese R. (2012). Il Finanziamento delle Politiche Regionali per la Non Autosufficienza in Puglia. Ipotesi di Riforma dell'Addizionale Irpef. In: Peragine V. (a cura di). *Povertà e Politiche in Puglia*. Roma: Carocci.

Ceriani L., Figari F. and Fiorio C. (2013). EUROMOD Country Report. Italy 2009-2012. Colchester: University of Essex. Available at: -- https://www.euromod.ac.uk/sites/default/files/country-reports/Year4/CR_IT2009-2012_FINAL.pdf.

Cipollone A., Coromaldi M., Curci N., De Luca G., Depalo D. and Rossetti C. (2013). The Non-Behavioral Module of the Italian Microsimulation Model Econlav. *Government of the Italian Republic (Italy), Ministry of Economy and Finance, Department of the Treasury Working Paper* 2.

Cozzolino M., Declich C., Di Marco M., Proto G., Solari F. and Tanda P. (2015). Modelling Social Security, Direct Taxes and Cash Benefits. *Rivista di statistica ufficiale*, 2.

Decancq K., Decoster A., Spiritus K. and Verbist G. (2012). MEFISTO: A New Microsimulation Model for Flanders. FLEMOSI Discussion Paper 14.

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

Di Nicola F., Mongelli G. and Pellegrino S. (2015). The Static Microsimulation Model of the Italian Department of Finance: Structure and First Results Regarding Income and Housing Taxation. *Economia pubblica*, 2: 125-157.

Eurostat (2016). EU statistics on income and living conditions (EU-SILC) methodology, -- http://ec.europa.eu/eurostat/statistics-explained/index.php/EU_statistics_on_income_and_living_conditions_(EU-SILC)_methodology, downloaded on 26 October 2016.

Fiorio C. (2009). Microsimulation and Analysis of Income Distribution: An Application in Italy. *VDM, Verlag*.

Gastaldi F., Pazienza M.G. and Pollastri C. (2017). I Modelli di Microsimulazione dell'UPB: Famiglie e Imprese. *mimeo*.

Judkins D.R. (1990). Fay's Method for Variance Estimation. *Journal of Official Statistics*, 6(3): 223-239.

Maitino M.L. and Sciclone N. (2008). Il Modello di Microsimulazione Multiregionale dell'Irpet Microreg. *SIEP Working Paper* 604.

Maitino M.L., Ravagli L. and Sciclone N. (2013). Il Modello di Microsimulazione dell'Irpet MicroReg: Aggiornamenti ed Estensioni. *Studi e Approfondimenti, IRPET*. Available at: -- http://www. irpet.it/storage/pubblicazioneallegato/438MICROREGfinale. pdf.

Mitton L., Sutherland H. and Weeks M. (2000). *Microsimulation Modelling for Policy Analysis*. Cambridge: Cambridge University Press.

Morciano M. (2006). Gli Effetti Redistributivi dei Tributi Comunali di Puglia ed Emilia Romagna. *Economia Pubblica*, 3: 87-110.

Pellegrino S., Piacenza M. and Turati G. (2011). Developing a Static Microsimulation Model for the Analysis of Housing Taxation in Italy. *International Journal of Microsimulation*, 4(2): 73-85.

Popova D. (2012). Constructing the Tax-Benefit Micro Simulation Model for Russia-RUSMOD. *EUROMOD Working Paper* 7.
Proto G. (1999). Il Modello di Microsimulazione MASTRICT: Struttura e Risultati. *Rivista di statistica ufficiale*, 3.
Rizzi D. (1990). Un Modello Integrato per lo Studio degli Effetti delle Politiche Economiche sulla Poverta. *Politica Economica*, 2.
Silverman B.W. (1986). *Density Estimation for Statistics and Data Analysis.* CRC press, 26.
Sutherland H. (1991). Constructing a Tax-Benefit Model: What Advice Can One Give? *Review of Income and Wealth*, 37(2): 199-219.
Sutherland H. and Figari F. (2013). EUROMOD: the European Union Tax-Benefit Microsimulation Model. *International Journal of Microsimulation*, 1(6): 4-26.
Wilkinson K. (2009). Adapting EUROMOD for Use in a Developing Country: The case of South Africa and SAMOD. EUROMOD Working Paper 5.
Žarković-Rakić J. (2010). First Serbian Tax-Benefit Microsimulation Model-SRMOD. *The Quarterly Monitor of Economic Trends and Policies in Serbia*, 20.